

<https://orcid.org/0009-0007-3070-3355>
<https://www.cienciavita.pt/pt/9410-7EBE-7C48>

Computational Simulation Models in Typeface Legibility Research

*Assessing the Reliability of the VI-OCR Pipeline as an
Alternative to Tests with Visually Impaired Participants*

Author: Hugo Carvalho
Master's Supervisor: Professor Pedro Manuel Reis Amado, Ph.D.
Dissertation submitted in partial fulfillment of the requirements for
the Master's Degree in Graphic Design and Editorial Projects.
FBAUP, June, 2026

<https://orcid.org/0009-0007-3070-3355>
<https://www.cienciavita.pt/pt/9410-7EBE-7C48>

Modelos de Simulação Computacional na Investigação da Legibilidade de Tipos de Letra

*Avaliação da Fiabilidade da Pipeline VI-OCR como
Alternativa a Testes a Participantes com Deficiência
Visual*

Autor: Hugo Carvalho
Orientador: Professor Pedro Manuel Reis Amado, Ph.D.
Dissertação apresentada para cumprimento parcial dos requisitos
necessários à obtenção do grau de Mestre em Design Gráfico
e Projetos Editoriais.
FBAUP, junho, 2026

Acknowledgements

This thesis is the result of a journey that has been as much about personal discovery as it has been about type design and research.

I am deeply grateful to those who stood by me, challenged me, and opened doors that I did not even know existed. To my family, especially to my parents, Cristina and Miguel, I would like to thank you for your unconditional belief in me. You gave me a world where I could dream without limits and the space to grow into exactly who I wanted to be. Everything I achieve is built on the foundation you created.

To my best friend, Vicente, thank you for being there through it all. From the serious breakthroughs to the endless goofing around, you've been the best buddy anyone could ask for. As we've often discussed, I look forward to us growing old together and remaining friends until the end of our days.

To my colleagues in the MDGPE, thank you for being such an incredible and inspiring group to graduate with. I truly believe you are some of the best designers in the country. In particular, I want to thank Jorge Duarte, and Ana Coelho for helping on the execution of my test protocol, and Julien Lima for helping me with some initial technical issues and questions.

I would like to also extend my sincere thanks to ACAPO – Associação dos Cegos e Amblíopes de Portugal for welcoming my work and for the collaboration of everyone who participated in my test protocol, with a special thanks to Tiago Silva for his vital help in making this coordination possible, and also to Qingying Gao and Dr. Yingzi Xiong from the Johns Hopkins University for allowing me to conduct this research and accepting to collaborate throughout its development, without a doubt it is amazing to be a part of something much larger than myself.

A significant part of this research exists because of the generosity of the type design community. I am profoundly grateful to Morisawa, Dave Crossland, Dan Rhatigan, Erik Van Blokland, Mary Meürer, Fabian Harb, Ricardo Santos, Octavio Pardo, Maíra Woloszyn, Gerry Leonidas, Erik Spiekermann, Aprígio Morgado, and James Edmondson for their time and insights. They say you shouldn't meet your heroes, but in this field, the opposite is true. To the rest of the type design community, meeting so many of you at Fontstand 2026 in Berlin was a turning point for me, the kindness and openness of this community are rare, and I feel lucky to be part of it.

Finally, my deepest thanks to my supervisor, Professor Pedro Amado. You were the one who first sparked my fascination with type design and research. I knew from the first time we met, a few years ago, that I wanted to learn with and be mentored by you. Thank you for pushing me to surpass my own limits, for “nerding out” with me about several topics and for being the exact person I needed to find my way in this field.

Abstract

This study presents the results of a comparative investigation into the effectiveness of typeface legibility testing methodologies for people with visual impairments. The field of accessible type design faces significant challenges, as evaluating the legibility of fonts has traditionally been a complex and time-consuming process that relies on testing with human participants. The rise of technologies such as the “VI-OCR” Optical Character Recognition pipeline and human retina simulation models offers an opportunity to streamline and optimize this process. The primary objective of this research was to determine if simulation technology can serve as a reliable typeface legibility testing method, thereby evaluating if and how a pipeline such as VI-OCR can serve as an efficient substitute for human experimentation.

To achieve this goal, the study applied the testing method used by Larson in the development of the Sitka typeface, a validated technique for assessing individual letter recognition, to different fonts. The quasi-experimental methodology was divided into two parallel approaches. In the current research, legibility was first analyzed through a VI-OCR pipeline and a computational retina model, simulating different types and levels of visual impairment. Concurrently, the same tests were conducted with a group of human participants with similar various types of visual impairments. The results demonstrated a close link between the data obtained from the technological simulations and the results from the human participant tests, confirming the hypothesis that the technology could accurately and consistently mimic human visual perception.

The main conclusion affirmed that retina simulation models are powerful tools for legibility research, offering an approach that is faster, more consistent, and less prone to fatigue than traditional human testing. The importance of this study resides in its capacity to provide a proven methodology for aiding the development of accessible type design, allowing designers and researchers to evaluate legibility in a more agile, cost-effective, and systematic manner. This advancement can accelerate the development of fonts and research that can significantly improve accessibility and the reading experience for people with different types of visual impairments.

Keywords: Accessibility, Type Design, Visual Impairment, Impairment Simulation, Font Legibility, Legibility Testing, Optical Character Recognition, Vision Science, Computational Technology, Computational Design, Typefaces, Inclusive Design.

Resumo

Este estudo apresenta os resultados de uma investigação comparativa sobre a eficácia das metodologias de teste de legibilidade de tipos de letra para pessoas com deficiência visual. O campo do design de tipos de letra acessíveis enfrenta desafios significativos, uma vez que a avaliação da legibilidade de fontes tem sido tradicionalmente um processo complexo e demorado que depende de testes com participantes humanos. O aparecimento de tecnologias como o pipeline de Reconhecimento Ótico de Caracteres “VI-OCR” e modelos de simulação da retina humana oferece uma oportunidade para simplificar e otimizar este processo. O principal objetivo desta investigação foi determinar se a tecnologia de simulação pode servir como um método fiável de testar a legibilidade de tipos de letra, avaliando assim se, e como um pipeline como o VI-OCR pode servir como um substituto eficiente para a experimentação humana.

Para alcançar este objetivo, o estudo aplicou a diferentes fontes o método de teste utilizado por Larson no desenvolvimento do tipo de letra Sitka, uma técnica validada para avaliar o reconhecimento de letras individuais. A metodologia quasi-experimental foi dividida em duas abordagens paralelas. Na presente investigação, a legibilidade foi analisada primeiro através de um pipeline VI-OCR e de um modelo computacional de retina, simulando diferentes tipos e níveis de deficiência visual. Em simultâneo, os mesmos testes foram realizados com um grupo de participantes humanos com vários tipos semelhantes de deficiência visual. Os resultados demonstraram uma forte ligação entre os dados obtidos a partir das simulações tecnológicas e os resultados dos testes com participantes humanos, confirmando a hipótese de que a tecnologia poderia simular de forma precisa e consistente a percepção visual humana.

A principal conclusão afirmou que os modelos de simulação de retina são ferramentas poderosas para a investigação de legibilidade, oferecendo uma abordagem que é mais rápida, mais consistente e menos propensa à fadiga do que os testes humanos tradicionais. A importância deste estudo reside na sua capacidade de fornecer uma metodologia comprovada para apoiar o desenvolvimento de design de tipos de letra acessíveis, permitindo a designers e investigadores avaliar a legibilidade de uma forma mais ágil, económica e sistemática. Este avanço pode acelerar o desenvolvimento de fontes e investigações que podem melhorar significativamente a acessibilidade e a experiência de leitura para pessoas com diferentes tipos de deficiência visual.

Palavras-chave: Acessibilidade, Design de Tipos de Letra, Deficiência Visual, Simulação de Deficiência, Legibilidade de Fontes, Testes de Legibilidade, Reconhecimento Ótico de Caracteres, Ciência da Visão, Tecnologia Computacional, Design Computacional, Tipos de Letra, Design Inclusivo.

Table of Contents

Acknowledgements	3
Abstract	4
Resumo	5
Table of Contents	6
List of Figures	8
Statement of Authorship	10
I. Introduction	11
I.I. Context	12
I.II. Research Problem (and hypothesis)	12
I.III. Research and Operational Goals	13
I.IV. Key Concepts	14
I.V. Significance of the Study and Motivations (Intrinsic and extrinsic motivations)	15
I.VI. Research Methodology	16
I.VII. Study Delimitations	18
I.VIII. Document Structure	19
II. Literature Review	20
II.I. Adopted Type of Literature Review	21
II.II. Areas of Research	21
1. Introduction to Type Design	24
1.1. Defining Type Design: Communication and System	26
1.2. Foundational Typographic Terminology	27
1.3. Designing for Inclusion: The Importance of Accessible Typography	28
2. The Process and Practice of Type Design	30
2.1. The Modern Type Workflow: From Concept to Digital Font	32
2.2. Design Motivation and Context: Targeting Specific Reading Needs	36
2.3. Typographic Parameters as Variables for Legibility Testing	38
3. Legibility: Concepts, Human Factor and Assessment	42
3.1. The Biomechanics of Reading: Visual and Perceptual Constraints	44
3.2. Impact of Visual Impairments on Typeface Legibility	47
3.3. Contemporary Research on Improving Legibility	48
3.4. Typeface Legibility Assessment Methods	50
4. Computational Tools and Models for Legibility	53
4.1. Foundational Models of Visual Degradation Simulation	55
4.2. Precursors and Alternative Comput. Legibility Assessment Approaches	59
4.3. Optical Character Recognition	61
5. VI-OCR and its Application in Type Design	63
5.1. The VI-OCR Framework: Architecture and Logic	65
5.2. Clinical Parameterization and Retinal Simulation	67
5.3. OCR as a Proxy for Human Recognition Patterns	68
5.4. The Utility of VI-OCR in the Type Design Process	69
Literature Review Summary	71
III. Empirical Work	73
6. Methodology	74
6.1. Methodological Approach	75
6.2. Research Site	77
6.3. Participants and Sampling	78
6.4. Data Collection Instruments	81
6.5. Ethical Considerations	84
6.6. Operationalization	84
7. Data Presentation and Findings	94
7.1. Results: Description and Analysis	95
7.1.1. Primary Research Question	100
7.1.2. Secondary Research Questions	104
7.2. Discussion of Findings	106
7.3. Study Limitations, Constraints and Adopted Contingency Measures	112

IV. Conclusion	114
8. Summary of Results and Deliverables	115
9. Discussion of Results	118
10. Dissemination	120
11. Limitations, Improvements and Recommendations for Future Work	122
V. End-Matter	127
1. References	128
2. Appendices	135
Appendix A - Contingency Plan	136
Appendix B - Supplementary Material - Terminology	137
Appendix B - Supplementary Material - Typographic Anatomy	139
Appendix B - Supplementary Material - Distribution/Publication	141
Appendix C - Stimuli (Original and Filtered)	145
Appendix D - Informed Consent Forms	146
Appendix E - Test Protocol Data	153
Appendix F - Questionnaires/Interviews	154

List of Figures

I. Introduction	11
Fig.1. Methodology Diagram.	17
II. Literature Review	20
Fig.2. Classification of severity of vision impairment.	22
Fig.3. Ellen Lupton's Type Classification system.	32
Fig.4. The process of grid-fitting and hinting.	34
Fig.5. Motivations across several types of organization scales.	36
Fig.6. Advantages of vision simulation models over human testing.	38
Fig.7. Serif positioning in Tiphares, following Beier and Dyson's study.	39
Fig.8. Sans Forgetica.	40
Fig. 9. Atkinson Hyperlegible.	40
Fig.10. Bionic Reading.	41
Fig.11. Distribution of Fixation Durations During Reading.	44
Fig.12. Word shape recognition.	45
Fig.13. Word shape recognition using the envelope around the word.	45
Fig.14. Cone of Vision.	46
Fig.15. Why Helvetica sucks at small sizes, especially for user interfaces.	46
Fig.16. Coarse-to-fine sequence of spatial frequency processing.	55
Fig.17. Eye Disease Simulations.	56
Fig.18. Visual effects of PSF.	57
Fig.19. Effect of Eccentricity on Letter Recognition.	58
Fig.20. Demonstrating and mapping the spatial limits of crowding.	59
Fig.21. Slides from IJUP 2026, explaining the VI-OCR pipeline.	65
Fig.22. Slides from IJUP 2026, explaining the VI-OCR pipeline.	66
Fig.23. VI-OCR Simulator with Full Pipeline.	67
III. Empirical Work	73
Fig.24. Comparison between each typeface.	76
Fig.25. Example of Test Protocol Room Layout.	78
Fig.26. Example of presented Stimuli from the Test Protocol.	79
Fig.27. Summary of the Test Protocol.	81
Fig.28. Reading Material from the "Jornal de Notícias" Newspaper.	82
Fig.29. Reading Material from the National Geographic Magazine.	82
Fig.30. Signed Consent Form.	85
Fig.31. Precision Vision ETDRS Chart.	86
Fig.32. Precision Vision Pelli-Robson Contrast Sensitivity Chart.	87
Fig.33. Example of used Noise Image to clear participants vision.	88
Fig.34. VI-OCR Pipeline Environment.	89
Fig.35. VI-OCR Pipeline Environment 2.	89
Fig.36. VI-OCR Pipeline Environment 3.	90
Fig.37. VI-OCR Pipeline Environment 4.	90
Fig.38. VI-OCR Outputs folder.	91
Fig.39. VI-OCR Example of .Json file OCR results for VI-OCR01.	91
Fig.40. Raw Test Protocol Data.	92
Fig.41. Table showcasing the results from the Staircase Procedure.	95
Fig.42. Edited Tables showcasing the Results per each typeface.	96
Fig.43. Example of mathematical analysis between P02 and VI-OCR.	97
Fig.44. Comparison of the performance for each of the test typefaces.	99
Fig.45. Kappa results for each participant and VI-OCR.	100
Fig.46. P-Value results for each participant and VI-OCR.	101
Fig.47. Relationship Between Kappa and P-Value.	102
Fig.48. Performance of each typeface.	104
Fig.49. Comparison between MAE and MSD.	105
IV. Conclusion	114
Fig.50. Website developed in the context of VI-OCR's original research	123
Fig.51. Iterations of our web-based VI-OCR implementation.	123

V. End-Matter	127
Fig.52. Constituent parts of letters.	139
Fig.53. Scientific Results.	140
Fig.54. Findings from the 2025 Fontstand Survey.	143
Fig.55. GT Mechanik Microsite.	143

Statement of Authorship

I hereby declare that the contents of this document are of my own authorship and have not been previously used in another course or curriculum unit, at this or another institution. References to authors (statements, ideas, thoughts) meticulously comply with the rules of attribution, and are duly indicated in the text and bibliographical references, in accordance with referencing standards. I am aware the practice of plagiarism and self-plagiarism constitutes academic misconduct.

I hereby declare that, in the present document, artificial intelligence tools were utilized. DeepSeek V3.2 was used to assist with debugging, modifying, and fixing the VI-OCR pipeline, Google Gemini 2.5 Flash and OpenAI GPT-5 were used to edit, analyse and proofread content. This content was properly curated, evaluated and verified by me. The formatting, in-text citations, and reference list of this thesis have been prepared in accordance with the standards of the American Psychological Association (7th ed.), commonly referred to as APA 7. This style governs the structural organization of the manuscript, the presentation of data, and the ethical attribution of all scholarly sources cited throughout the research.

A handwritten signature in black ink, appearing to read 'Jorge Carrillo', written in a cursive style.

(UPorto, 2026)

I.

Introduction

I.I. Context

The present study is positioned at the intersection of the fields of digital accessibility, type design, cognitive science and computational technology. This research is primarily driven by the importance of inclusive design, a subject that has gained considerable momentum in recent years.

From a social perspective, and in spite of growing awareness, it directly responds to a global urgency to develop more accessible, enhanced reading experiences for the substantial population of low-vision users (WHO, 2019).

This is particularly critical because current typeface development is often insufficient to meet the specific physiological needs of these individuals.

Additionally, this study aims to explore an alternative to traditional methods for testing the legibility of typefaces. As Beier (2012) highlights, these methods are fundamentally constrained, with human-subject testing being often characterized as logistically complex, resource-intensive and scientifically variable.

Whilst human-to-human interaction is a logical part of any scientific procedure, and the heart of development of new products, constraints, like participant fatigue, subject bias, shifting environmental conditions and high resource requirements prevent the systematic, large-scale evaluation of design parameters and subconsequent faster and easier development of accessible typefaces.

Technologically, the context of this work is defined by the recent convergence of advanced computer vision and vision science. Recent developments in the field have led to the creation of high-fidelity Computer Retinal Models (Dömötör & Horváth, 2025), which allow, in theory, for the precise simulation of pathological degradation. When integrated with modern Optical Character Recognition (OCR) systems that utilize artificial intelligence, to simulate the behavior of test subjects, these models offer a potentially transformative approach to legibility testing.

In alignment with ongoing international research conducted in partnership with Johns Hopkins University, this dissertation seeks to empirically implement and evaluate a modified version of the VI-OCR pipeline originally proposed by Gao et al. (2025). With this approach being proposed as a reliable and agile way to simulate visually impaired vision, with the potential, as previously mentioned, to significantly accelerate the development of evidence-based, accessible typefaces.

Crucially, this study was not developed in isolation; Its relevance is, of course, rooted in its potential ability to resonate with type design professionals and academic researchers. To ensure the research remained grounded in the realities of the industry I engaged directly with the global type design community, participating in the 2026 Fontstand Conference in Berlin, maintaining active involvement across key industry channels, and conducting a series of expert inquiries.

I.II. Research Problem (and hypothesis)

Testing typeface legibility, particularly for specialized reader populations such as those with visual impairments, has historically been tethered to extensive, resource-intensive human-subject experimentation. As documented by Beier (2012), this traditional approach is fundamentally constrained by logistical complexities, specifically regarding the specialized recruitment of participants with various visual pathologies (WHO, 2019). Furthermore, as previously mentioned, such methods are frequently compromised by high variability, subjective environmental factors, and participant fatigue. Rendering the

systematic and iterative testing of numerous design parameters, such as the vast parametric space of a Variable Font, impractical or even impossible for academic researchers and smaller type foundries. This opens up the opportunity to develop and test an agile, and systematic alternative to traditional human experimentation.

As highlighted, this technological promise leads directly to the central, unresolved research problem:

Can computational simulation models, specifically an adapted VI-OCR (Visually Impaired Optical Character Recognition) pipeline (Gao et al., 2025), serve as reliable and functionally equivalent method of testing typeface legibility with visually impaired participants, thereby accelerating the development of accessible typefaces?

This research aims to close this gap by directly comparing the quantitative outcomes of the computational model against human testing results, following a protocol similar to the one documented by Larson in the development of the Sitka typeface (Larson and Carter, 2016), utilizing trigrams and a staircase procedure. The validation of such a tool, as mentioned before, is essential to democratize and accelerate evidence-based accessible type design.

I.III. Research and Operational Goals

This study seeks to demonstrate that the VI-OCR pipeline proposed by Gao et al (2025) can accurately and consistently replicate the perceptual outcomes of human readers under compromised viewing conditions, thereby accelerating the design cycle for evidence-based accessible typefaces.

To achieve this primary goal, the research is organized into several operational objectives, which serve as necessary facilitators for the final validation:

1. Implementation of a rigorous testing protocol that directly compares the quantitative output of the VI-OCR pipeline against the data derived from human-subject testing, this protocol will be developed following accepted legibility testing procedures, in this case utilizing trigrams and a staircase procedure (Larson & Carter, 2016). This, consequently, also involves analyzing how different typefaces behave within both the simulation and the human-subject data sets.
2. Assess the practical feasibility of the computational tool by personally executing the VI-OCR pipeline across multiple tests. This stage aims to deconstruct the “black box” of the simulation to understand its internal mechanics and identify the technical requirements for its deployment in a design environment.
3. Evaluate the professional receptivity and market fit of the VI-OCR pipeline within the type design industry by administering structured, open-ended written questionnaires to a sample of expert type designers and researchers. These qualitative inquiries aim to gather industry perspectives on market demand, isolate critical ethical and methodological themes regarding AI-driven simulation, and map practical integration pathways into existing studio design workflows.

In view of these objectives, the central Research Question is:

To what extent do the legibility metrics generated by the VI-OCR computational simulation pipeline reliably correlate with data obtained from human participants using a test protocol based on Larson’s research (Larson & Carter, 2016)?

From what has been previously discussed, we also establish the following secondary research questions:

1. How accurately can the VI-OCR pipeline replicate specific visual pathologies?
2. What are the underlying mechanics that dictate how AI algorithms process the OCR pass on the degraded image of different typeface letter expressions?
3. How do different typefaces and specific design parameters behave when subjected to the same computational degradation protocol compared to human vision impairments?
4. How feasible is the integration of the VI-OCR pipeline into the standard creative workflow of a type design studio?
5. What is the learning curve for designers attempting to run these simulations independently?
6. Is there demand within the type design industry for computational legibility tools?
7. What are the primary barriers to the adoption of such a protocol in professional practice?

This study is based on the hypothesis that the specific character recognition errors made by the VI-OCR pipeline will strongly correlate with the errors made by human participants. It is anticipated that the simulation will accurately simulate impaired human vision, under specific testing conditions, confirming its capacity as a predictive tool. Furthermore, it is proposed that the professional community will, in general, identify a high utility for this tool, provided the pipeline remains accessible, ethical, and technically manageable for smaller studios.

I.IV. Key Concepts

This section provides a preliminary definition of the core concepts and variables that underpin this research. These definitions serve as the basis for the empirical testing phase and will be expanded upon in the subsequent literature review section.

Computer Retinal Models (CRM): In the context of this study, CRM refers to the mathematical and computational simulation of the human eye's optical and neural degradation. This technology, used and described by researchers like Dömötör and Horváth (2025), allows for the theoretical precise replication of the human eye and visual conditions.

VI-OCR Pipeline: This is defined, according to Gao et al. (2025), as the integrated system consisting of a degradation model (CRM) and an Artificial Intelligence recognition engine, using OCR libraries like Tesseract or AI models, such as ChatGPT, Gemini, and others. It serves as the primary technical pipeline for assessing the accessibility of typefaces without initial human intervention.

The visually impaired: Individuals, which retain some usable vision, the primary target for accessible typography, as opposed to individuals relying on non-visual methods like Braille or audio transcription, for example. The WHO (2019) classifies the individuals relevant to this study from having Mild Visual to Severe Visual Impairment (visual acuity in the better eye: 6/12 - 6/60).

To facilitate the empirical validation of the VI-OCR pipeline, this study identifies a clear relationship between independent and dependent variables.

The experimental design utilizes three true Independent Variables: Typeface Design (categorical; Sitka, Luciole, Arial, and Times New Roman), Typographic Scale (ordinal; systematically manipulated from 120pt down to 24pt), and the Visual Degradation Profile (operationalized biologically via participant visual pathologies in the human group, and algorithmically via calibrated Contrast Sensitivity Function variables in the machine group). The foundational Dependent Variable across all trials is Character Recognition Accuracy, measured as a binary outcome (Pass/Fail). Within the broader scope of the study, this setup is analyzed via a Concurrent Predictive Validation Framework: the human legibility performance serves as the validation criteria against which the accuracy of the simulations created by the VI-OCR computational simulation pipeline are evaluated.

Consequently, it is hypothesized that a strong, positive correlation exists between the Visual Degradation Profiles (the primary independent variable) and the resulting legibility outcomes (the dependent variable) across both testing environments. More specifically, it is anticipated that the predictive data set generated by the VI-OCR pipeline and the data set obtained from human subjects will demonstrate parallel failure patterns under matched degradation conditions. Ultimately, demonstrating this consistent relationship across both groups is expected to validate that the simulation can reliably approximate human visual perception for specific profiles of visual impairment.

I.V. Significance of the Study and Motivations (Intrinsic and extrinsic motivations)

The significance of this research lies at the intersection of social responsibility, technological innovation, and the evolution of type design methodologies. As inclusive design gains global momentum, there is an urgent need for tools that bridge the gap between academic vision science and the practical requirements of type foundries.

My primary intrinsic motivation is rooted in a deep-seated interest in the synergy between type design, technology, accessibility, inclusion and research. This study offers a personal opportunity to contribute to type design research and to help improve legibility and accessibility for people with visual impairments. By engaging directly with global experts and established research institutions, this work is positioned within the contemporary discourse of inclusive design, driven by the personal goal of improving the daily reading experiences of the visually impaired population.

From an extrinsic perspective, as previously mentioned this study addresses a critical gap regarding the limitations of traditional legibility testing procedures. By contributing to on-going international research and the development of VI-OCR, the potential to have a pipeline that allows type designers and researchers to conduct systematic testing is incredibly valuable, opening opportunities such as, for example, extensive testing of the full parametric space of a variable typeface. The primary extrinsic motivation, therefore, is to validate the VI-OCR computational simulation pipeline (Gao et al., 2025) as a reliable and agile alternative for initial human testing.

Beyond this, providing a cost-effective simulation pipeline can enable type designers to easily conduct typeface legibility testing, identifying and fixing legibility issues early in the creative process, leading to more accessible typefaces on the market for the visually impaired.

In summary, this research is not only a response to a personal academic interest but a necessary step towards modernizing the type design industry's approach to inclusion. By establishing a proven pipeline, this study aims to bridge the gap between theoretical vision models and the practical realities of the global type community.

I.VI. Research Methodology

The nature of this study is fundamentally Quantitative, employing an Quasi-Experimental Design. This methodology is defined by the objective necessity of comparing quantifiable error rates and confusion matrices. The research is structured in a quasi-experimental manner to systematically measure the dependent variable across two distinct independent processes, as previously described, one utilizing human participants, and the other utilizing the VI-OCR computational model pipeline (Gao et al., 2025), in order to establish a statistical correlation between them.

With this quantitative and quasi-experimental design in mind (Figure 1), the research is structured into four core stages: Theoretical Research, Pre-Testing, Parallel Data Collection and Crossing, analysis and Results.

As previously mentioned the success of this research hinges on establishing a statistical correlation between human and machine recognition errors. This requires the rigorous control of variables and the meticulous execution of parallel testing procedures.

Human testing is conducted on a targeted convenience sample of Visually Impaired (VI) voluntary participants, recruited together with the ACAPO association, who are fluent in the Latin script and of Portuguese origin. These individuals are the definitive users for whom accessible typography is designed (as previously described, with visual acuity in the better eye between 6/12 - 6/60). The testing follows an adapted version of the testing procedure conducted by Larson & Carter (2016).

As described by Larson and Carter (2016), in the original testing procedure participants are presented with short, three-letter non-words (to prevent guessing via reading) at brief display times. A staircase procedure dynamically adjusts the presentation time for each participant until they reach a predetermined 50% accuracy rate. This ensures participants are operating at the threshold of recognition, which is essential for capturing the character-to-character confusion patterns. In the testing procedure for this study an adaptation was made, as visually impaired people need much longer times to recognize letters and in the case of VI-OCR testing, time is not a measurable, nor relevant metric. This led to the decision of opting to alternatively run an adaptation of the staircase procedure, with trigrams varying in size from 120pt to 24pt in four different typefaces.

The computational counterpart, as previously mentioned, uses the VI-OCR (Visually Impaired Optical Character Recognition) Pipeline (Gao et al., 2025), to objectively mimic the human perceptual process under low-vision constraints. The identical set of trigram images (the three-letter non-words) used in the human tests is input into the VI-OCR pipeline. The pipeline's first step applies a Computer Retinal Model filter. This filter is designed using Contrast Sensitivity Functions (CSFs) that are precisely parameterized to match the average visual acuity (VA) and contrast sensitivity (CS) loss of the VI human participants sample. To match these values and degrade the image to simulate the vision of the participants, the filtering is impacted by several different metrics such as Hshift, Vshift, Monitor Distance, Monitor Resolution (pixels) and Monitor size (Inches).

This output image is then processed by a high-performance Optical Character Recognition (OCR) engine utilizing AI models such as Google Gemini, ChatGPT, Tesseract, etc.

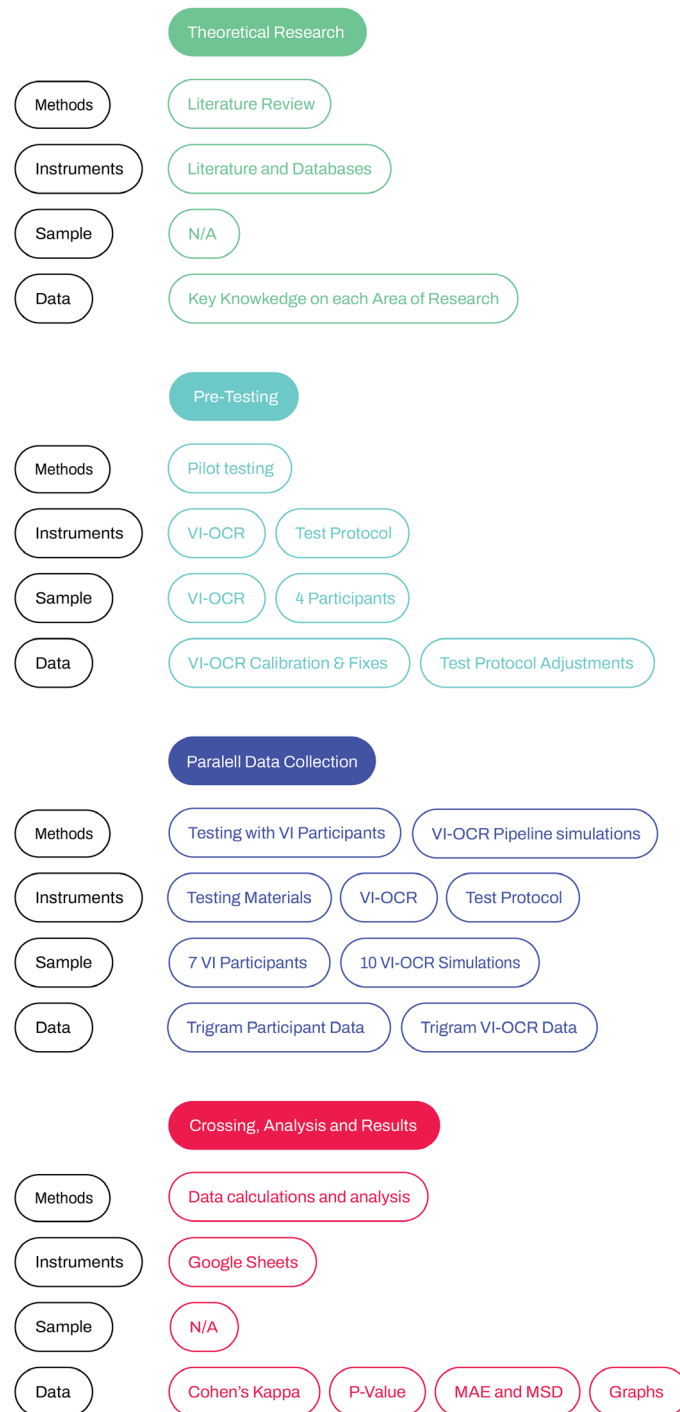


Fig.1. Methodology Diagram.

The final validation is achieved by calculating the statistical correlation between the Human Error Rates and the Computational Error Rates, thereby confirming the predictive validity of the VI-OCR model.

As this type of research requires extensive knowledge in fields outside of the scope of the Master's Degree curriculum. A contingency plan (see Appendix A) was developed in case we wouldn't be able to set up and run VI-OCR, which wouldn't allow us to move past the "Pre-testing" stage as seen on Figure 1.

Although this will be explained later, in higher detail, we were initially unsuccessful in setting up the repository of VI-OCR created by Qingying Gao and Dr.Yingzi Xiong, opting instead to push forward with a web-based implementation that was developed based on their paper. After a few months of progress, the original implementation from the repository was fixed and we were able to run it successfully, and the web-based implementation was not further developed.

This study is designed to achieve two closely linked sets of results: scientific outcomes, which resolve the research problem, and applied outcomes, which deliver practical value to the design community.

The primary scientific result is the statistical validation of the VI-OCR pipeline, which will be confirmed by a robust correlation coefficient between the Computational Error Rates generated by the tool and the Human Error Rates collected in testing. This successful validation proves that the automated simulation can reliably predict human character recognition performance under low-vision constraints, establishing its credibility as a research tool. Furthermore, the study will yield a set of quantified confusion matrices, detailing the precise patterns of error, and establish evidence-based metrics by quantifying the impact of specific typographic parameters on legibility performance.

In the applied domain, the most tangible outcome is the creation of a typeface legibility testing methodology. The validated VI-OCR pipeline provides an objective, efficient complementary tool for the costly and time-consuming process of human-subject testing. This tool will allow typeface designers to rapidly iterate and optimize typeface candidates for Low Vision readers, with small studios and researchers benefiting the most.

I.VII. Study Delimitations

Unfortunately, as this study is part of a Master’s program it is impossible to test every situation and condition, hence why it was important to establish the delimitations of this study. This allowed for the research to still answer our questions, fulfill the goals we set out to achieve and be relevant in the context of the ongoing international research, whilst remaining manageable for the time frame we’re given to develop our study.

Within this context, the following delimitations were established:

Does not test all possible typefaces or parameters

This dissertation does not aim to explore the full typographic design space or optimize every possible typeface or Variable Font parameter. It only analyzes a selected set of typographic features for a selected group of typefaces and their relationship to VI-OCR and the user-testing outcomes, because a full exploration would be unmanageable within the scope and time of a master’s thesis.

Does not replace human testing entirely

The research does not claim to fully replace human-subject testing or to validate VI-OCR for all future legibility studies. It only tests whether the pipeline can act as an initial pre-validation filter that correlates with one specific, established user-testing method, and one specific test group, since proving universal validity across all tests, populations and conditions would require a much larger, multi-study program.

Limited to one type of model and one human test method

This dissertation does not compare multiple computational models or multiple human legibility protocols. It focuses exclusively on one simulation approach (VI-OCR pipeline) and one human test protocol in order to keep the study methodologically coherent and technically feasible.

Does not cover all visual impairments or user contexts

This dissertation does not attempt to model every type and severity of visual impairment, nor all reading contexts (e.g., motion, glare, complex layouts). It only examines conditions that can be reasonably simulated in the VI-OCR model and matched in the experimental human testing, to ensure controlled comparison.

Does not control for precise clinical variations in the participant sample.

This dissertation does not utilize a clinically controlled or stratified sample of participants with visual impairments. Because the study relies on a participant sample gathered through a collaboration with ACAPO, the research trusts and defers to the organization's internal recruitment to assemble a group meeting our general criteria. Consequently, it is not possible to strictly control, independently measure, or standardize the exact nature, severity, or stability of each participant's visual impairment within the scope of this Master's thesis.

I.VIII. Document Structure

This dissertation is organized into a four-chapter structure designed to transition the study from theoretical validation to empirical proof. Chapter 1, the Introduction, establishes the social and technological context, identifies the core research problem (the need for a validated alternative to human testing), defines the research objectives, and presents the central hypothesis. Chapter 2, the Literature Review, systematically builds the foundation for the research, integrating four distinct thematic areas: an Introduction to Type Design (practice and process), Legibility (concepts, human factors, and assessment methods), Computational Tools and Models for Legibility (detailing CSF, PSF, retinal simulation, and other relevant concepts), and finally, VI-OCR and its Application in Type Design (which is the subject of testing). Chapter 3, presents the empirical study, encompassing the Methodology and Procedures of the quantitative, quasi-experimental design, which involves parallel data collection from human participants and the VI-OCR pipeline. This chapter also includes the Analysis and Discussion of the resulting statistical correlation and confusion matrices. Finally, Chapter 4, the Conclusion, will provide an overview of the research outcomes, assess the validation of VI-OCR as an agile design optimization tool, detail the project's deliverables, and propose avenues for future research.

II.

Literature Review

¹ *Morisawa is a globally renowned Japanese font company, widely considered a leader in developing high-quality, technically sophisticated CJK typefaces, making their input invaluable (see Appendices for questionnaire details).*

II.I. Adopted Type of Literature Review

The methodology adopted for this initial review is mainly Thematic Synthesis. In future chapters, mostly those where simulation and computation models are the central focal point, an exploratory approach is going to be used. By choosing these approaches I was able to systematically group existing research around a few core conceptual dimensions: Type Design; Accessibility; Legibility Assessment in type, and Computational Simulation Models. This structure allows for a critical evaluation of established knowledge within each domain, enabling the clear isolation and justification of the specific research that this proposal aims to address.

II.II. Areas of Research

As previously mentioned, the literature review is organized around a few core and interconnected concepts, establishing the necessary theoretical and methodological context for the validation study:

1. Type Design

This topic of research establishes the systematic nature of typeface design, grounded in both aesthetic and functional theory. Foundational literature defines the craft by its historical evolution and structural principles, with this in mind core books such as *The Stroke Theory* (Gerrit Noordzij, 1985/2005), *Theory of Type Design* (Unger, 2018) and *Thinking with type: a critical guide for designers, writers, editors & students* (Lupton, 2010) were invaluable sources that helped to contextualize and frame the core concepts. Authors like Karen Cheng (2005) are central in defining anatomy and the specific properties (eg. cap height, ascenders, descenders, counters, etc) that comprise the visual system of a typeface.

For the purpose of this study, the scope is deliberately limited to Latin typefaces. This decision was made by taking into consideration the unique complexity and different visual systems of Cyrillic, Greek, Arabic, CJK, Indic, or other complex scripts or minority languages, whose legibility requires distinct research methodologies not within the current scope of this study. Furthermore, the choice of limiting this study to only Latin typefaces is a necessary consideration for the human testing process, ensuring that the participants in the study are highly familiar with the script, thereby isolating legibility challenges due to visual impairments rather than unfamiliarity or cognitive load.

Although, as previously mentioned, the decision was made to strictly limit the research to the Latin script, there remained a significant personal interest in understanding diverse visual systems, and their legibility relationships. To gain insight into these challenges, particularly concerning the relationship between character density and visual impairment, a questionnaire was developed and submitted to the type designers at Morisawa¹. Within the Latin script context, research by Sofie Beier (2012) confirms the value of designing for quantitative parameters, demonstrating how specific features like x-height and aperture are key variables for improving legibility in low-acuity vision (Cooreman & Beier, 2024). The emergence of Variable Fonts (Maíra Woloszyn, 2022) is particularly relevant here, as this technology provides the essential parametric system for generating the large volume of test candidates needed for systematic, evidence-based legibility analysis.

2. Accessibility

In the following topic of research, we transition from the technical structure of type design to its functional application in meeting diverse human needs. Accessibility in the context of this study focuses on designing for human diver-

sity (Kat Holmes, 2018) to ensure that the written word is functionally usable for individuals with Visual Impairments (VI), particularly those categorized with Low Vision (World Health Organization, 2019). As referenced previously, these individuals, who retain some usable vision, are the primary target for accessible typography, and the focus of our study group. In opposition individuals relying on non-visual methods like Braille or audio transcription, for example, are not relevant for this research. According to the classification table below (Figure 2), we can categorize the relevant participants of this study as those classified in between the Mild Visual Impairment to Severe Visual Impairment groups (visual acuity in the better eye between 6/12 - 6/60).

Category		Visual acuity in the better eye	
		Worse than:	Equal to or better than:
Mild vision impairment		6/12	6/18
Moderate vision impairment		6/18	6/60
Severe vision impairment		6/60	3/60
Blindness		3/60	
Near vision impairment		N6 or M 0.8 at 40cm	

Fig. 2. Classification of severity of vision impairment based on visual acuity in the better eye [Table]. In World Health Organization (Ed.), World Report On Vision (2019).

This focus necessitates a deep dive into Human Factors in reading. The process of character recognition is governed by visual and perceptual constraints, involving rapid fixations and saccades (eye movements) that rely on the acuity of foveal vision and the context provided by peripheral vision (Sofie Beier, 2012). Visual impairments, such as Macular Degeneration or Glaucoma, compromise these mechanisms. This degradation leads to critical perceptual issues like the crowding effect, where characters become visually indistinguishable due to insufficient separation (Sofie Beier, 2012; Sofie Beier et al., 2022). This understanding justifies the need for compensatory typographic design, specifically manipulating parameters like Set Width and Tracking to ensure character distinctiveness under compromised reading conditions (Sofie Beier, 2012). Research into type design legibility, such as the development of the Sitka typeface, confirms the importance of testing individual letters in context (e.g., three-letter sequences) to measure the impact of neighboring characters (Larson & Carter, 2016).

The final consideration in this thematic area is the reading medium used for testing, specifically visual displays. This study focuses on assessing the performance of VI-OCR against human testing, whilst analyzing the behavior of several typefaces, which can be of course utilized across both print and digital media. The test protocol was developed with these considerations in mind, although, as mentioned before on the study delimitations, it is impossible to test every condition, every typeface and every environmental change. Standardized legibility norms, such as ISO 24509:2019 (ISO, 2019), already address factors like viewing distance, luminance, and contrast to estimate the minimum legible font size for people with no pathological disorders. However,

this ISO standard doesn't address people with visual impairments in its scope, reinforcing the need for research specifically focused on VI. To achieve this focus and maintain the scientific purity necessary for validation, this research, as previously mentioned, will not extend to the analysis of variable conditions of the reading environment (e.g., ambient light, screen glare) or the specific effects of digital rendering techniques (e.g., anti-aliasing methods). This approach aligns with methodologies used in legibility tests, where typefaces are tested on high-resolution displays at scaled distances to avoid digital artifacts and focus solely on the geometric performance of the letterforms (Larson & Carter, 2016).

3. Legibility Assessment in type

This dimension focuses on the established science of legibility assessment. Research confirms that character recognition is fundamentally dependent on human visual processes, which are compromised by visual impairments (Sofie Beier et al., 2022). For assessment of legibility the field of type design usually relies on extensive human-subject testing using proven research, here it is important to define what these proven research methods are and why the decision was made to implement the same type of testing protocol as Larson, which, as noted before, informed the development of typefaces like Sitka (Larson & Carter, 2016). This type of protocol is versatile and easy to adapt to a simulation of the visual conditions of a real human-being, whilst yielding the relevant important data needed to cross with the data attained from the tests with the participants.

4. Computational Simulation Models

This final thematic area examines the technological solutions emerging from computer vision and machine learning to overcome the scientific and logistical limitations of human-subject testing (Sofie Beier, 2012). The objective of research in this field is to identify, analyze, and compare existing simulation pipelines and tools to understand their mechanical viability for automated legibility assessment. Related to this, it is crucial to understand two primary concepts, computer retinal models and optical character recognition (OCR), and how their intersection forms a broader methodology for legibility testing. Computer Retinal Models apply digital image processing techniques to simulate the visual degradation experienced by low-vision readers. This involves filtering the input image based on parameters like Visual Acuity (VA) and Contrast Sensitivity (CS), mimicking specific pathologies such as Macular Degeneration or Glaucoma (Dömötör & Horváth, 2025).

The resulting degraded image is subsequently processed by an OCR engine. Unlike traditional commercial applications where OCR is optimized to maximize text recognition accuracy, this methodological approach utilizes the OCR engine's error rates to mimic human reading mistakes. While specific pipelines like VI-OCR (Gao et al., 2025) formalize this into a script we can run, the underlying mechanism relies on the broader principle that an OCR's failure patterns under digital degradation can replicate human low-vision performance (Dömötör & Horváth, 2025).

The most promising result of this conceptual fusion is the Visually Impaired Optical Character Recognition (VI-OCR) pipeline which has been extensively mentioned. The VI-OCR pipeline (Gao et al., 2025) is specifically designed to bridge the gap between computer vision and the low-vision field, aiming to estimate whether text of a given size would be recognizable for a low-vision human reader. The core challenge lies not just in constructing the pipeline, but in demonstrating that its output, its automated error rate, is statistically correlated with the human confusion matrices (Larson & Carter, 2016).

1.

Introduction to Type Design

Chapter Summary:

→ This sub-chapter lays the foundation for understanding Type Design as a critical, systematic discipline that bridges visual culture, technology, and cognitive science, with the fundamental goal of optimizing written communication for the human reader. It defines the field through its core components and outlines the traditional, highly variable creative process of type design. A central focus of the chapter is drawing a precise distinction between readability, the holistic performance of reading continuous text, and legibility, defined as the ease of distinguishing individual characters based on inherent design qualities like stroke thickness and internal spacing. Crucially, it concludes by framing type design as an exercise in universal accessibility and human inclusion. It highlights the necessity of engineering accessible typefaces to maximize the remaining reading capability of individuals with Low Vision who navigate the specific visual challenges imposed by eye conditions like Glaucoma, Macular Degeneration, Cataracts, and Diabetic Retinopathy. Ultimately, this urgent need for accessible design is affected by the profound logistical, ethical, and human challenges of validating typographic solutions through traditional legibility testing methodologies.

1.1. Defining Type Design: Communication and System

To understand type design we first have to understand language and its importance for humanity. According to Cheng (2005), “language is arguably the most significant human characteristic. Through a complex system of signs, sounds and symbols, ideas and messages are communicated to our ever-expanding global society - information that has the power to shape and change civilization itself.”

Type Design is the specialized discipline concerned with the creation of complete systems of letterforms, punctuation, and symbols, known collectively as a typeface, intended for the efficient delivery of written communication. While often appreciated for its aesthetic qualities, the fundamental purpose of Type Design is functional: to optimize the visual structure of text for quick and accurate recognition by the human reader. It operates as a complex visual system where the success of the whole is contingent upon the harmonious relationships between individual components.

A critical distinction in this field, particularly relevant to digital and computational analysis, is that between the typeface and the font. The typeface refers to the abstract, underlying visual design, the style, characteristics, and inherent structure of the letterforms themselves. The font, by contrast, is the functional, implementable instance of that design, in modern digital contexts, this is the data file (such as OpenType or TrueType) that enables the rendering of the typeface across various mediums. In the context of this research, the typeface is the visual object of study, and the font is the digital structure that facilitates both human consumption and computational simulation.

The systematic nature of Type Design requires adherence to consistent parameters across the glyph set. This includes defining rules for stroke weights, character proportions, and, crucially, the internal and external spacing between glyphs. This systematic rigor is what transforms an alphabet into a tool for continuous reading flow (Beier, 2012), ensuring that the visual texture of the text supports, rather than hinders, the cognitive process of translating symbols into meaning. Therefore, Type Design is best understood as a critical interface between visual culture, technological implementation, and cognitive processing, making it a key domain for scientific research into reading efficiency and accessibility.

Despite its expressive potential, Type Design is fundamentally a discipline of precision and functional constraints. The necessity of adapting text to new environments has fueled rapid growth in the industry, leading to a significant increase in the number of available typefaces (Cheng, 2005). This expansion is largely a response to technological innovation. The appearance of new ways of consuming information, such as the computer, the internet and the smartphone has demanded that typefaces be designed with different purposes in mind. High-density digital displays and other technological implementations have demanded that these typefaces were instead designed as robust, technically optimized systems, rather than purely aesthetic objects.

In the same way, typefaces have also been designed with other specific purposes in mind, such as helping children who are just starting to learn how to read, the dyslexic and the visually impaired. The methodology for creating a typeface is highly variable, reflecting the unique nature of individual designers and the diverse motivations behind their work (Napoleão, 2021). The impulse to create a new font can stem from a variety of factors, such as, personal intellectual inquiry, market demand for specialized audiences, or a functional necessity (to overcome problematic viewing conditions, for example). Despite the lack of a single 'correct' process, type design practices tend to address specific typographic parameters as part of their development, though

the emphasis and sequence vary widely among designers. This typically begins with sketching a few key letters (such as the lowercase 'a', 'e', 'g', 'n', and 'o') to establish the core proportions and personality of the design. These initial forms are then tested using specific word sequences, such as the frequently selected term “*hamburgefontiv*”, before the full glyph set is fleshed out and ultimately digitized, a process requiring precise tracing and outlining to convert manual sketches into scalable vector forms utilizing specialized software such as Glyphs or Fontlab (Cheng, 2005). Finally, the typeface undergoes rigorous testing and quality assurance procedures to ensure the structural system meets its intended functional criteria.

1.2. Foundational Typographic Terminology

The field of Type Design relies on precise vocabulary to differentiate the various components of a letterform and the resulting effects on the reading experience. For this research, which centers on computational evaluation and the analysis of these components, it is essential to establish clear, functional definitions for the core terms related to the design object (See Appendix B). The distinction between the design and its implementation is foundational, especially in a digital context where the former is the visual subject and the latter is the data structure that facilitates simulation.

While often used interchangeably in general discourse, legibility and readability describe two distinct perceptual phenomena, and their differentiation is critical for designing experimental evaluation protocols.

Legibility refers to the ease and speed with which the human eye can distinguish one letter or character from another. It is primarily a function of the typeface’s inherent design qualities, such as character form, stroke thickness, aperture size, and internal character spacing (Beier, 2012). This research focuses heavily on legibility, as the foundational success of character recognition is a prerequisite for any higher-level reading process. High legibility is particularly vital in challenging viewing conditions, such as those experienced by individuals with visual impairments.

Readability refers to the user’s experience of reading continuous text. However, as noted by Beier et al. (2022), the term suffers from significant variation across disciplines. To clarify the position of this research, it is necessary to distinguish between three common interpretations.

Traditional designers often define readability as the result of “invisible” typography, affected by factors like line length, leading, and alignment (Warde, 1955; Lupton, 2010). In this view, readability is how well the layout facilitates the flow of text.

In psychology, readability is often linked to the reader’s comprehension and the linguistic complexity of the content (Rayner & Pollatsek, 1989). Here, a text is “readable” if the vocabulary and syntax are appropriate for the audience. Early researchers like Luckiesh and Moss (1939) and Tinker & Paterson (1932), measured readability through performance metrics, such as involuntary blink rates or speed of reading.

For the purposes of this dissertation, we adopt the definition proposed by Beier et al. (2022), which views readability as an interdisciplinary metric of performance (speed and accuracy) influenced by the interaction between the typeface (legibility), the layout (typographic variables), and the reader (vision and familiarity). While acknowledging that readability is the ultimate goal of a text system, this study isolates legibility as the primary variable to better understand how incremental design changes, enabled by Variable Font technology, impact the baseline of optical character recognition in VI-OCR.

Equally as important, the typographic features of a typeface or the “metrics” are the quantifiable parameters that govern the spacing and flow of text. These parameters are directly manipulable variables in type design that affect human perception and, consequentially, apply to all computational models aiming to replicate visual perception.

1.3. Designing for Inclusion: The Importance of Accessible Typography

The goal of type design extends beyond aesthetic appeal and general reading efficiency, it is fundamentally an exercise in universal accessibility. As the written word is the primary vehicle for information, education, and civic engagement, the ability to read efficiently is a foundational component of modern life. Designing for inclusion recognizes that the general population is not homogeneous (Holmes, 2018), particularly concerning visual function, which can be altered by various conditions leading to visual impairments.

According to Holmes (2018):

Designing for human diversity might be the key to our collective future. It is going to take a great diversity of talent, working together, to address the challenges we face in the 21st century: climate change, urbanization, mass migration, increased longevity and aging populations, early childhood development, social isolation, education, and caring for the most vulnerable among us in an ever-widening gap of economic disparity. You never know where, or who, a great solution will come from. (p. 12)

This insight necessitates a fundamental shift in how organizations, institutions, and governments operate. The priority should be to actively cultivate inclusive ecosystems not just for representation, but to ensure that the optimal solutions for complex global issues, from sustainable infrastructure to equitable healthcare, are found at the intersection of various lived experiences. This demands that we dismantle systemic barriers, empowering individuals from marginalized or non-traditional backgrounds not only to be present but to actively shape the design process and lead implementation efforts. Only then can we truly build a resilient and equitable future.

Taking this into consideration, understanding the population for whom accessible type is critical requires precise terminology:

Visually Impaired: People with partial or total vision loss that cannot be fully corrected with standard corrective lenses like glasses or contact lenses.

The Blind/Blindness: This term generally refers to those with severe visual impairment or a complete lack of light perception. According to the World Health Organization (2019), this is defined as visual acuity of less than 3/60 or a visual field less than 10 degrees in the better eye with the best possible correction. For this group, type design is often irrelevant, as they rely on non-visual reading methods (e.g., Braille, audio descriptions, or screen readers that convert text to speech).

Low Vision: This subcategory of VI refers to individuals who have some usable vision, often defined as visual acuity between 6/12 and 6/60 in the better eye, but whose vision cannot be fully corrected. This group is the primary beneficiary of accessible type design, as carefully designed typefaces can significantly enhance their remaining reading capability by maximizing the clarity and distinctiveness of each character (Beier et al., 2022).

Glaucoma: This is a group of eye conditions that damage the optic nerve, often due to abnormally high pressure in the eye. A key consequence is the grad-

ual loss of peripheral (side) vision, which can severely impact the visual field required to track lines of text and successfully integrate the foveal (central) fixations necessary for reading (Beier et al., 2022).

Macular Degeneration (MD): A common, age-related condition that causes damage to the macula (the center of the retina), resulting in a loss of central vision. Since the macula is responsible for sharp, detailed central vision, the exact area used for fixating on individual characters, MD is one of the most debilitating conditions for reading, often creating a central blind spot that forces the reader to use less acute peripheral vision.

Cataracts: A condition characterized by the clouding of the eye's natural lens. This results in generalized blurring, decreased contrast sensitivity, and an increased susceptibility to glare. For typographic design, this means that features like high contrast strokes (very thin parts) and tight letter spacing are often rendered visually indistinguishable due to the light scattering effects of the cloudy lens.

Diabetic Retinopathy (DR): A complication of diabetes caused by damage to the blood vessels of the light-sensitive tissue at the back of the eye (the retina). DR can lead to the creation of scotomas (small blind spots) or localized visual distortion across the field of vision, making continuous and predictable reading flow extremely difficult.

The widespread need for accessible typography is contrasted by the profound difficulty of testing and validating design solutions. Traditional legibility testing, the gold standard for validating accessible type design, is fraught with logistical, ethical, and methodological challenges.

Human experimentation is time-consuming, and subject to high variability due to the unique combination of visual, cognitive, and attentional differences among participants. Moreover, ensuring reproducibility is difficult: a participant's performance can be influenced by attention, motivation, and, critically, fatigue, as confirmed by existing research (Beier, 2012).

2. The Process and Practice of Type Design

Chapter Summary:

→ This sub-chapter establishes the modern typeface design workflow as a rigorous intersection of artistic expression, digital engineering, and functional problem-solving. Moving beyond traditional craft, the contemporary process is defined as an iterative system engineering cycle comprising five core stages: Conceptualization, Drafting, Production, Validation (QA), and Distribution. Central to this workflow is the transformation of aesthetic concepts into quantifiable typographic variables. Key parameters, such as x-height, set width, aperture openness, and stroke modulation, are identified not merely as stylistic choices but as functional levers that dictate legibility. The chapter emphasizes that for specialized contexts, particularly accessible design for the visually impaired (VI), these parameters must be meticulously engineered to compensate for physiological deficits like reduced contrast sensitivity and the crowding effect.

A significant critical focus is placed on the Validation phase, noting the structural and financial barriers that often prevent independent designers from conducting rigorous human-subject testing. To address this “resource gap,” the chapter proposes computational simulation models (such as VI-OCR) as a strategic methodological bridge. By utilizing Variable Font (VF) technology to isolate design axes, these simulations can potentially allow for an evidence-based, iterative testing loop that enables designers to refine typefaces for maximum performance before final deployment. Ultimately, the chapter positions the published typeface as a piece of digital communication infrastructure, where its success is measured by its objective performance across a diverse spectrum of human, environmental, and technological constraints.

2.1. The Modern Type Workflow: From Concept to Digital Font

The contemporary development of a typeface represents a rigorous intersection of art, engineering, and human-computer interaction. Moving decisively beyond the historical limitations of metal and photo typesetting, the modern workflow is best understood as a process of digital system engineering and intricate problem-solving (Freitas, 2017). This process is inherently iterative and non-linear, which means it isn't shaped by a rigid sequence but by the designer's personal motivations, the project's specific constraints, and the ultimate context of use (Napoleão, 2021). The central challenge remains converting fluid concepts and aesthetic choices into precise, quantifiable parameters capable of rendering legible text across diverse digital and editorial environments.

The complex process of creating typefaces can be broadly divided into five overlapping and interconnected phases: Conceptualization, Drafting and Design, Production and Engineering, Validation/QA, and Distribution/Publication. This five-stage model serves as a pedagogical and analytical framework to dissect the entire design cycle, moving from abstract intention to concrete, functional output (Napoleão, 2021; Freitas, 2017).

Conceptualization and Research

This initial phase establishes the purpose, personality, and technical requirements of the typeface. It is driven by the project brief, which can range from an aesthetic exploration to a highly specific functional requirement, such as designing for maximum legibility in high-ambient-light screen environments or for users with visual impairments. Research into historical precedents, linguistic needs, and intended media (print vs. screen) informs critical early decisions on the fundamental structure (Napoleão, 2021), including:

Aesthetic Style (eg.: Serif, sans-serif, slab, script, or display) (Figure 3).

SERIF	MINION	Aa	GEORGIA	Aa	BODONI	Aa
	HUMANIST		TRANSITIONAL		MODERN	
SLAB	TISA	Aa	CHAPARRAL	Aa	KULTURISTA	Aa
	HUMANIST SLAB SERIF		TRANSITIONAL SLAB		GEOMETRIC SLAB	
SANS	GILL SANS	Aa	HELVETICA	Aa	FUTURA	Aa
	HUMANIST SANS SERIF		TRANSITIONAL SANS		GEOMETRIC SANS SERIF	

Fig.3. Ellen Lupton’s Type Classification system. Adapted from Lupton (2010), as cited in Amado (2022)

Writing System Support: Latin, Cyrillic, Greek, Arabic, CJK, Indic, or other complex scripts or minority languages.

Core Metrics: eg. Establishing the x-height for latin based scripts, and the overall stroke weight, contrast (modulation) (eg. chinese/japanese) based on functional and stylistic goals.

Drafting and Design

This phase involves the physical shaping of the letters, typically starting with a small set of “control” characters (such as n, o, H, O, l, and g) that define the typeface’s overall rhythm and proportional logic (Freitas, 2017).

Digital Drafting with Bézier Curves: The forms are translated into digital outlines using Bézier curve (mathematical curves defined by anchor points and control handles). This conversion is crucial as it dictates the scalability and technical performance of the typeface. Mastery lies in minimizing the number of points while maintaining smooth, optically correct outlines.

Interpolation and Masters: For type families with multiple weights and widths or other options defined in the conceptualization and research, the designer utilizes master designs (e.g., Light and Bold). Interpolation software then mathematically generates intermediate weights (Regular, Semibold) between these masters. This system, while foundational, is the precursor to the more advanced control offered by Variable Fonts.

Production and Engineering

This is the technical heart of the workflow where the font’s behavior is defined. The effort shifts from purely formal design to metric and code refinement (Freitas, 2017). In this phase, the typeface is transformed from a static collection of glyphs into a functional text engine through several critical engineering processes:

Advanced Layout Logic: This involves scripting the OpenType features that govern character behavior, such as contextual alternates (calt) for script connectivity, discretionary ligatures (dlig), and stylistic sets (ss01-ss20) that allow for aesthetic flexibility without compromising the core character set (Peter-Con, 2022).

Diacritic and Accent Engineering: The systematic construction of localized glyphs using anchor points. This ensures that marks (e.g., tildes, cedillas) are positioned with mathematical precision relative to the base character, a process vital for the “highly-engineered” nature of modern typefaces.

Font Info and Global Metrics: Defining the foundational metadata of the file, including vertical metrics (ascender, descender, and x-height values), Panose classification values for system-level font matching, and naming tables that ensure cross-platform compatibility.

Variable Design Space Mapping: If utilizing variable technology, this includes defining the axis mapping and interpolation rules. This ensures that as a user moves a slider, for instance, between a Light and Bold weight, the transition remains visually consistent and technically stable.

Horizontal Metrics and the “Drawing” of Space

While often categorized as a post-production task, the definition of horizontal metrics is fundamentally an extension of the drawing process. The designer does not merely create a glyph shape; they define the set width, the total horizontal footprint of the character, by balancing the “black” of the stroke with the “white” of the side bearings (Noordzij, 1985/2005). **Integrated Spacing:** This “drawing of space” creates the primary rhythm of the typeface. In modern font editors like Glyphs or FontLab, this is managed through spacing strings (e.g., “nnnnn” or “ooooo”), where the designer adjusts side bearings to ensure the internal character counters match the external space between characters (Glyphs GmbH, 2024; Fontlab Ltd, 2026).

Kerning as Refinement: Beyond the basic side bearings, kerning acts as a corrective layer for specific character pairs (e.g., To, VA) or character classes (where similar shapes like o, ó, ö share the same spacing rules). This process ensures an “optically even texture,” preventing visual gaps that disrupt the saccadic movements of the eye during reading (Beier, 2012; Rayner & Pollatsek, 1989).

The Text Engine: The quality of these metrics determines the visual rhythm and the “tracking” (global letter-spacing) of the font. As Freitas (2017) notes, these metrics transform individual glyphs into a cohesive flow of continuous text, effectively “automating” the visual rhythm for the end user.

Hinting and Optimization: For non-Variable Fonts, hinting involves adding instructions to ensure clarity at low resolutions (Figure 4).

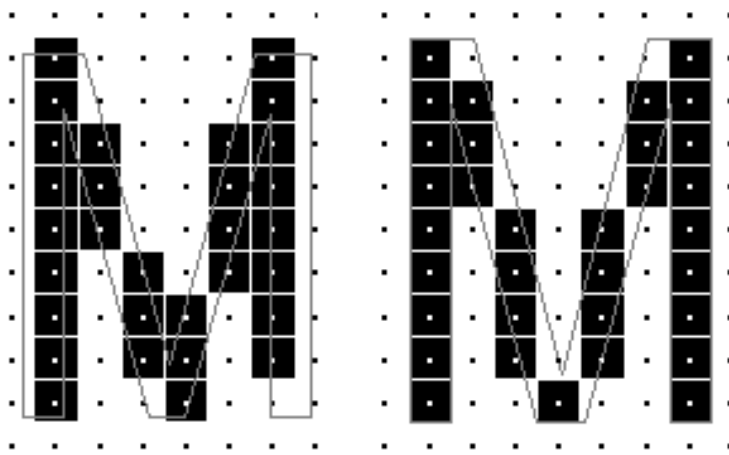


Fig.4. The process of grid-fitting and hinting to align vector outlines with pixel boundaries. Microsoft (2022).

Assessing the Quality of Typefaces

This stage involves assessing the font’s performance across target environments and against functional criteria, such as, for example, aesthetic balance, technical compliance, and legibility. Usually to ensure the developed typeface meets the established criteria checklists (Sousa, 2025) and heuristics can be used.

According to Leonidas (2015), the evaluation of a typeface must be grounded in the concept of “minimum quality”, a baseline of functional reliability that ensures a font remains legible and technically stable across all intended platforms. This involves assessing the “performance” of the design, where the designer must prove that the typeface maintains its identity and clarity under the stressors of low-resolution rendering or varied lighting conditions. In this context, this represents a transition from subjectivity to objective performance metrics, where the typeface is treated as a piece of infrastructure that must facilitate uninterrupted reading (Leonidas, 2013; Unger, 2018).

As reported by Sousa (2025) foundries like R-Typography utilize, as previously mentioned, a checklist to ensure quality of the final developed product. This type of framework involves verifying that the x-height proportions, stroke-to-counter ratios, and vertical metrics are consistent across the entire character set. This ensures that the “blackness” and density of the font remain uniform and that other problems are noted and fixed before the final release.

A critical component of modern QA is the validation of the OpenType layout logic. This involves a deep-level audit of the GSUB (Glyph Substitution) and GPOS (Glyph Positioning) tables (Hudson, 2018; Microsoft, 2024). Features such as contextual alternates, small caps, and discretionary ligatures must be tested for cross-environment compatibility. As noted in OpenType Specification (v1.9.1.), this type of engineering extends to the stability of the design space, the interpolation rules must be verified to ensure that no “outline collisions” or mathematical errors occur when the font is manipulated along its defined axes.

Finally, recent methodologies adopt the testing protocols seen in industry-standard pipelines, such as those utilized by Google Fonts, making use of automated technical auditing toolsets, like Fontbakery to identify “out-of-bounds” errors in naming tables, path directions, anchor point placement, and other critical problems of the design (Google Fonts, n.d.).

For accessibility-focused projects, this phase is non-negotiable, requiring proof that the design meets its functional objectives (Beier, 2012).

Although we have previously seen cases where legibility is part of the development of typefaces, such as Sitka, Frutiger, Tiphares, and Atkinson Hyperlegible the capacity to execute the Validation phase rigorously is directly dictated by the size and resources of the design entity, creating significant discrepancies in the pursuit of evidence-based design.

Consequently, smaller foundries and independent type designers who lack the financial resources to spend on legibility testing, either opt to not do it or rely on the use of informal methods that typically include reviews, subjective self-correction, or limited “hallway testing” with non-specialist peers. While these provide immediate feedback, they lack the empirical weight required for high-stakes legibility design. As noted by Leonidas (2013) and Beier (2012), subjective preference is a poor proxy for functional performance; a designer may “feel” a typeface is legible without accounting for the diverse physiological constraints of the end-user, such as low visual acuity or high-sensitivity to glare.

It is in the context of these resource-constrained environments (small studios, agencies with limited budgets, and academic projects) that computational simulation technologies find their core strategic utility for typeface design, not as a full alternative to human testing, but rather as another useful complementary tool that allows the accurate simulation of visual impairments, allowing designers to do initial testing of font legibility at no (or very low) extra cost.

Distribution/Publication

The modern landscape of font distribution is characterized by a systemic tension between centralized aggregators and decentralized, independent delivery systems. From some of the conversations I conducted at the 2026 Fontstand Conference in Berlin it was fairly easy to understand the current landscape of distribution and publishing in the type design industry. Large-scale marketplaces such as MyFonts (Monotype) function as the industry’s primary “warehouses,” leveraging sophisticated SEO algorithms and global brand recognition to dictate market visibility. These types of marketplaces usually present some advantages as they take care of a lot of the backend of the business, specifically regarding licensing and other processes. However, these platforms often impose a “curatorial” pressure that prioritizes commercial trends, whilst also taking a significant cut out of the foundry’s profit margins (about 50%), leading to a lot of foundries opting to self-publish. Some type designers instead opt for a different approach submitting and releasing their typefaces through already existing foundries, with certain ones, like for example, TypeTogether, going as far as offering scholarships providing mentorship, publishing, and

funding for new type designers (See more about Distribution/Publication on Appendix B).

2.2. Design Motivation and Context: Targeting Specific Reading Needs

Today, more than ever, we live in a world where there's a necessity to create products that fulfill more than just artistic expression. Functionality, accessibility, and social inclusion (Beier, 2012; Holmes, 2018) are more and more substantial reasons that move companies to innovate, adapt and push forward new products into the market. This transition positions the creation of type not just as a craft, but as an act of digital system design aimed at eliminating barriers to information access for a visually diverse population (Napoleão, 2021). This motivation is fundamentally rooted in the recognition that legibility is not a monolithic concept, but a spectrum dictated by individual visual, cognitive, and environmental conditions.

This model of contrasting motivations is synthesized on Figure 5 from the contextual drivers discussed in Napoleão (2021) and the functional goals outlined in Woloszyn (2022).

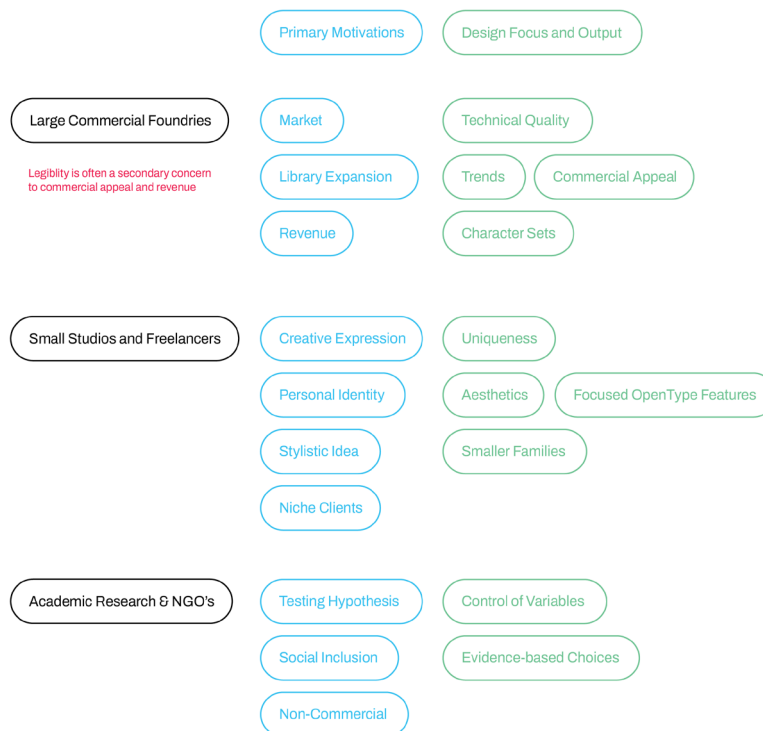


Fig. 5. Motivations across several types of organization scales.

Napoleão (2021) highlights that the “personal project”, driven by a designer's aesthetic curiosity, is a common starting point, but its subsequent development is immediately conditioned by its potential application, leading the designer to ask on what type of support will this be used (Napoleão, 2021).

These external factors define the three crucial contextual layers influencing the design:

Context of Identity and Culture

This context focuses on the typeface's role in conveying cultural meaning and reinforcing identity. The motivation here is to embed a sense of cultural identity or local heritage into the letterforms, often by referencing architectural details, historical calligraphy, or regional aesthetics (Freitas, 2017). This context heavily influences the aesthetic style (serif/sans, contrast), but its functional requirements are secondary to its expressive role.

Context of Media and Environment

This layer dictates the technical engineering requirements based on the final reading medium. The challenge here is optimization for the display technology and ambient reading conditions.

Print (Static Media): Requires stability and fine-tuning for physical processes (e.g., managing ink spread), focusing on detailed kerning and high-resolution output.

Screen (Dynamic Media): Demands robustness at low resolutions, optimization for anti-aliasing, and, critically, adaptability to varying user settings and ambient light. This is where the Grade axis in Variable Fonts becomes essential for adjusting density without reflowing the text.

Context of Function and User Need and Constraints

This context represents the highest functional requirement, where the design is motivated by a specific user deficit or a challenging environment (e.g., warning signs, automotive dashboards). In this realm, the design must overcome physical or cognitive barriers to ensure comprehension.

When the motivation becomes social inclusion, the design challenge changes from what looks good to what performs optimally under compromised reading conditions. This involves a deep understanding of human factors and the precise technical adjustments needed to compensate for various visual deficits.

Conditions like macular degeneration or diabetic retinopathy lead to reduced contrast sensitivity and blurring, necessitating increased stroke weight, larger x-heights, and maximized counter spaces (Beier, 2012; Beier et al., 2022). Conditions like glaucoma or retinitis pigmentosa cause peripheral vision loss, amplifying the crowding effect and demanding greater separation between characters (tracking) to allow for sequential recognition (Beier, 2012).

This commitment to accessible design imposes a crucial, non-negotiable requirement: the design must be rigorously validated to prove its superior functional efficacy. The challenge lies in the fact that the very target audience being designed for is the most difficult to recruit, test, and analyze using traditional methods.

Designing for specific reading needs necessitates robust validation, but, as stated earlier, the traditional human-subject legibility testing methodologies are often rendered impractical not just by financial burden, but by deeper structural and scientific limitations.

The analysis present on Figure 6 demonstrates that even in scenarios where the financial cost of recruiting subjects is mitigated (such as through collaboration with an NGO or academic funding), the limitations of time, logistics, and scientific purity remain significant barriers to efficient, evidence-based design.

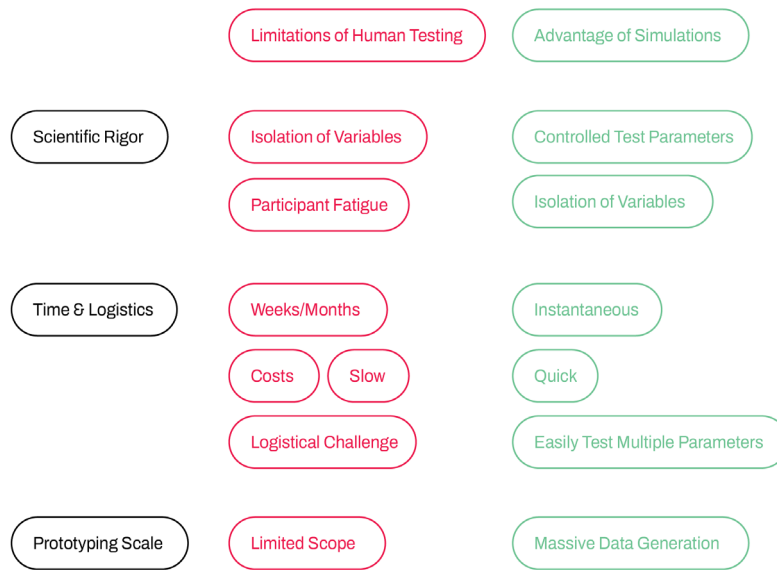


Fig. 6. Advantages of vision simulation models over human testing.

Therefore, computational simulation pipelines such as VI-OCR become essential for maintaining scientific rigor and enabling evidence-based design within all resource-constrained or time-sensitive environments, ensuring that the critical pursuit of accessible typography is not limited by the bottlenecks of traditional validation. An example of a possible methodological approach would be to use the simulation as a filter, allowing designers to confidently narrow the infinite possibilities of the Variable Font's design space down to the most promising candidates for final human verification.

2.3. Typographic Parameters as Variables for Legibility Testing

With static typefaces, testing involves looking at the specific masters that were developed or creating entirely new masters if you want to test a set range of the aforementioned properties (x-height, stroke weight, etc). Variable font technology has the potential to be extremely useful for this purpose, as you can develop and conduct testing along specific design axes. Although in this study we only utilized a specific weight of each tested typeface because of our set delimitations, it is important to understand the implications of Variable fonts for testing.

Research by Beier (2012) points to the importance of the height of lowercase letters, excluding ascenders and descenders. This being one of the most important parameters for perceived size and legibility. A larger x-height is crucial for users with reduced visual acuity, as it maximizes the distinct recognition area compensating for blurring and reduced resolution requires increasing the x-height to maximize the central reading zone. While a large x-height is good for local recognition, preserving sufficient ascender/descender length is vital for whole-word shape recognition, which aids fluent reading. The ratio between the cap height and x-height is often manipulated in accessible design to ensure capital letters are distinct but do not overwhelm the lowercase flow.

Increasing the set width provides crucial lateral separation between characters, which is a primary compensatory measure for visual field restriction and the crowding effect (Beier, 2012). Improper kerning can create disruptive visual gaps or crowding. Testing must ensure that adjustments to set width do not compromise the integrity of specific kerned pairs (Freitas, 2017). While set

width offers fine control, tracking provides a global variable to immediately increase the white space between letters across the entire text, offering immediate legibility gains at the cost of density (Beier, 2012).

Wide-open apertures prevent character forms from “closing up” or blurring into indistinguishable shapes (e.g., c becoming o) under low contrast or visual blur (Beier, 2012). Maximized counter spaces prevent blurring and are key to maintaining the unique identity of each letterform when the eye is only perceiving general masses (Beier, 2012). Minimal modulation (low contrast, closer to a monolinear structure) often performs better on screen and for low-acuity reading, as it maintains greater overall visibility of the stroke (Beier, 2012). Simple, easily distinguishable terminals (e.g., flat cuts over balls) contribute to clarity, while clear joins prevent strokes from merging visually (Beier, 2012).

Finally, the Weight axis is the most commonly recognized parameter, and its manipulation is crucial for controlling the visibility and contrast of the text against its background. Increasing the stroke weight directly improves contrast and visibility for readers with reduced contrast sensitivity, a common symptom of aging and various visual impairments. However, if increased too much, it can cause the counter spaces to shrink, leading to the visual fusion of adjacent letters, which is a phenomenon known as “clogging” or “merging.” This defines a critical upper limit for the Weight variable (Beier, 2012). A variable weight axis allows the designer to change the stroke weight and color density without changing the advance width (set width) of the characters (Woloszyn, 2022). This allows users to increase visibility (weight/grade) without causing text to reflow on the screen, making it a powerful tool for dynamic accessibility on digital platforms.

The refinement of these metrics is the “technical heart” of the workflow (Freitas, 2017), transforming a set of glyphs into a functional engine that controls reading speed and comprehension.

During the inquiries and interviews for the development of this dissertation I got to meet Octavio Pardo through an introduction made by Dave Crossland, not only did I find Octavio’s PhD research extremely interesting and his body of work at Ashler Design amazing, I found his development of the Tiphares typeface (released through his Pathfinders type foundry) extremely interesting to mention in the context of this topic.



Fig.7. Serif positioning in Tiphares, following Beier and Dyson’s study almost to the letter, except for the i and the j. Pardo (2025). <https://pathfinders.rocks/blog/the-path-to-tiphares-04-the-knowledge-behind-tiphares/>

The main reason I think it is an interesting project to mention is that it is a perfect example of a typeface built with functionality in mind, being consequently worried about how the typographic parameters affect its perception. Also because, compared to other typefaces, such as, Sans Forgetica (Figure 8), Atkinson Hyperlegible (Figure 9) and Bionic Reading (Figure 10) it doesn’t make any big claims that it is incredibly legible and solves all the problems in the world with little to no scientific backing behind it.

“The ‘desirable difficulty’ you experience when reading information formatted in Sans Forgetica prompts your brain to engage in deeper processing.”

Fig. 8. Sans Forgetica. Brinson (2019). <https://www.sambrinson.com/sans-forgetica/>

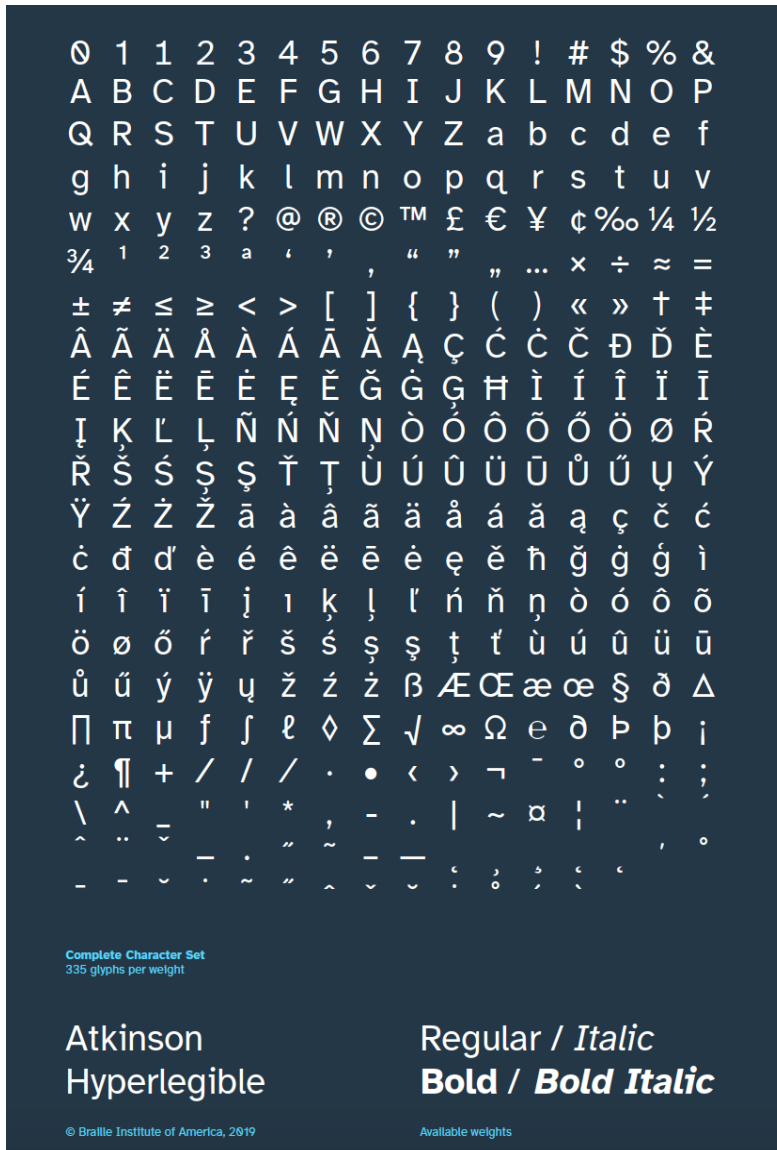


Fig. 9. Atkinson Hyperlegible. Braille Institute (2023). <https://www.brailleinstitute.org/freefont/>

Bionic Reading is a new method facilitating the reading process by guiding the eyes through text with artificial fixation points. As a result, the reader is only focusing on the highlighted initial letters and lets the brain center complete the word. In a digital world dominated by shallow forms of reading, Bionic Reading aims to encourage a more in-depth reading and understanding of written content.

Fig.10. Bionic Reading. Casutt, R. (2022, July 21). <https://bionic-reading.com>

In the blog posts explaining the development of Tiphares Pardo (2025) himself states that “it hasn’t been tested in scientific studies—so there’s no data, for or against, to claim that it performs better than Fago, Helvetica, or Frutiger”. This is where I think VI-OCR, or other Vision Computational Simulation Models, if proven as reliable, can be very useful pipelines allowing researchers like Pardo to conduct extensive initial testing to their typefaces, get initial conclusions to refine the design and then run final test protocols with the intended test group.

3.

Legibility:

Concepts, Human Factor
and Assessment

Chapter Summary:

→ This sub-chapter establishes that typeface legibility is a product of the interaction between type design and the biomechanical constraints of the human visual system. Reading is defined as a discontinuous cycle of rapid saccades and pauses known as fixations, during which the brain processes high-acuity information. Moving away from the historically popular “word-shape” theory, contemporary cognitive science supports a letter-recognition model, asserting that words are identified by first decoding their individual character components. This shift necessitates a design focus on character distinctiveness and the elimination of the confusion matrix among visually similar letters, as the burden of legibility rests on the efficiency of character-level encoding within the limited quarter-second window of a fixation.

The study of accessible typography is presented as a methodology of compensatory design aimed at mitigating the physiological failures of the eye. Pathologies such as cataracts, glaucoma, and macular degeneration degrade legibility by reducing contrast sensitivity, inducing visual blur, and amplifying the crowding effect. These conditions cause the “collapse” of negative space, where internal apertures fill in and adjacent letterforms merge into an indistinguishable texture. Consequently, the research identifies horizontal spacing and internal counterforms as the most critical variables for the visually impaired reader. By systematically increasing tracking and maximizing apertures, designers can move letterforms outside the expanded crowding zone of the impaired eye, transforming the typeface into a specialized tool for visual compensation.

The assessment of these design interventions requires a transition from subjective preference to objective, performance-based metrics. Rigorous legibility testing utilizes standardized protocols, such as Words Per Minute (WPM) measurements, LogMAR scaling for acuity, and the analysis of misrecognition matrices to identify structural weaknesses in the character set. While human-subject testing remains the gold standard, the chapter concludes by positioning computational simulation models, such as the VI-OCR pipeline, as an essential predictive tool. These simulations allow for the rapid and low-cost quantification of a typeface's resilience under stress, provided they can accurately emulate human recognition errors rather than achieving perfect machine recognition. This multifaceted approach ensures that the final design is grounded in empirical evidence and functional reliability.

3.1. The Biomechanics of Reading: Visual and Perceptual Constraints

The fundamental constraint on text legibility is the mechanical and neurological system of the human eye. Reading is not a smooth, continuous scan, but a highly precise, discontinuous cycle established by two primary eye movements: the saccade and the fixation (Rayner & Pollatsek, 1989). A saccade is a rapid, ballistic jump that moves the eye across the line of text, typically lasting only 20 to 30 milliseconds. During this movement, the reader experiences saccadic suppression, a functional blindness, meaning no information is absorbed. Fixations are the subsequent momentary pauses, lasting on average between 200 and 250 milliseconds (Figure 11), during which high-acuity visual information is captured and processed by the brain (Ferreira, 2014; Rayner, Slattery & Bélanger, 2010).

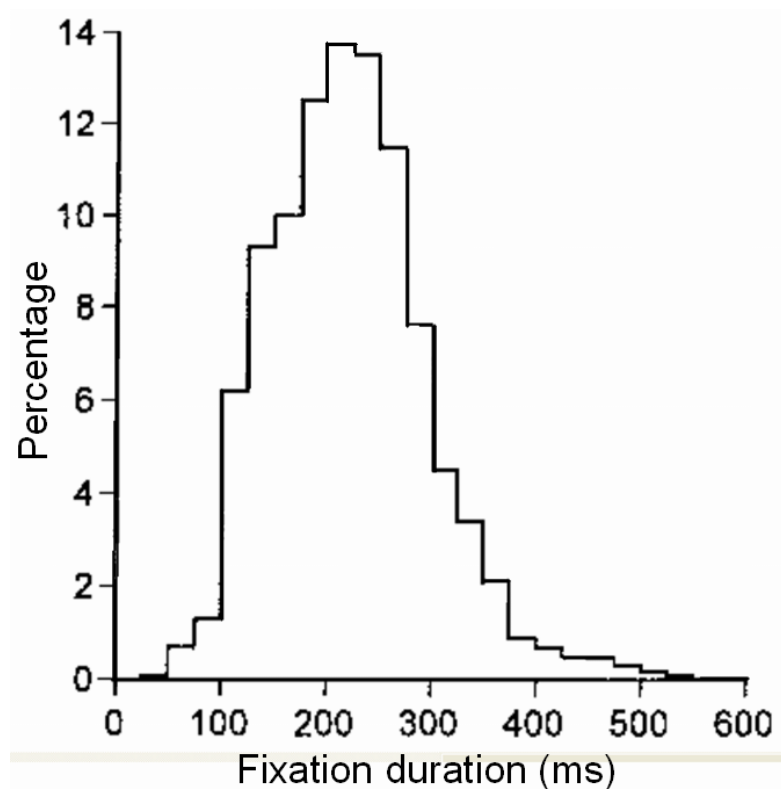


Fig. 11. Distribution of Fixation Durations During Reading. Rayner, K. (1998).

Pioneering psychological research, such as that by Cattell (1886), first highlighted that the time required for cerebral operations during reading is intrinsically linked to these fixation pauses. The overall efficiency of reading is therefore determined by the duration of the fixations (how quickly the information is processed) and the length of the saccades (how many characters are absorbed in each jump) (O'Regan, 1979). A more legible typeface is one that requires both shorter and fewer fixations to decode the same amount of text (Luckiesh & Moss, 1939). Crucially, the typographic design must be optimized to maximize the information captured in this brief quarter-second window, making the reading process less effortful and fatiguing.

The question of how the visual system recognizes a word has fueled typographic research for decades, generating a critical divergence between historical design belief and contemporary cognitive science.

1. This is the high-acuity center of the retina, responsible for precisely processing the two or three letters directly under the point of fixation. This is where definitive recognition occurs.

2. This is the critical surrounding area, extending approximately 10 to 15 characters to the right in Latin script (Rayner, Slattery & Belanger, 2010). The information gathered here is lower resolution but is essential for pre-processing. The reader uses the parafoveal input to program the trajectory and distance of the next saccadic movement, making it crucial for maintaining reading fluency. If this parafoveal information is degraded, the reader may execute a regressive saccade (jumping backward) or make a shorter, less efficient jump forward, severely reducing reading speed. This mechanism underscores why consistent and predictable inter-word spacing is essential.

The historically dominant theory in graphic design, often associated with figures like Gill (1931/2013) and Warde (1955), was the word-shape or “bouma” theory (Figure 12 & Figure 13). This model proposed that readers recognize common words holistically, by decoding their overall outline or unique contour, which is primarily defined by the pattern of ascenders, descenders, and x-heights. The word-shape model implies that the internal design of individual letters is secondary to the external silhouette of the word. Proponents of this theory often advocated for the use of Roman and serif types over geometric sans-serifs, arguing that the variety in the former’s letterforms produced a more distinct and recognizable outline, thereby facilitating speedier recognition. This holistic view supported the idea that, in the words of Warde (1955), typography should be “invisible” - meaning the reader should see the meaning, not the forms.



Fig.12. Word shape recognition using the pattern of ascending, descending, and neutral characters. Larson (2004).

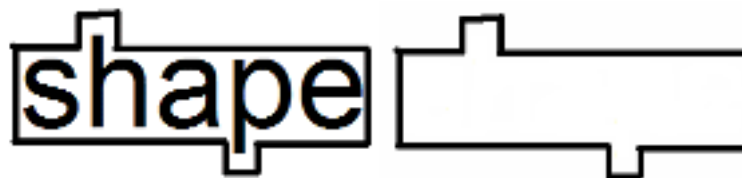
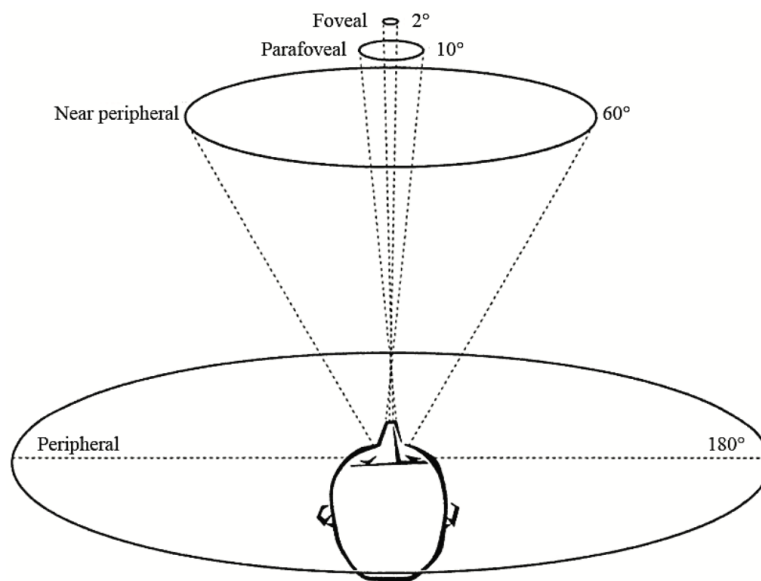


Fig.13. Word shape recognition using the envelope around the word. Larson (2004).

However, contemporary cognitive psychology and typeface legibility testing have largely refuted the word-shape theory, supporting the letter recognition model. This model asserts that words are recognized by first identifying their individual letter components, which are then assembled into the final lexical unit (Larson, 2004; Rayner et al., 2012). Research, particularly in psycholinguistics (McClelland & Johnston, 1977), demonstrates that reading efficiency is directly tied to the distinctiveness of the inner, not outer, parts of the character.

Larson (2004), in his work at Microsoft, forcefully argued this point: “We recognize words by first recognizing individual letters, then using the letters to build a word”. This statement which would later be applied in the development of the Sitka typeface (Larson & Carter, 2016) places the burden of legibility squarely on the design of the individual letterform, especially the distinctions between easily confused characters (e.g., lowercase i, l, and 1, or a, o, and c). This scientific consensus mandates that typographic decisions must prioritize clarity and differentiation at the letter-level to enable efficient encoding within the limited time frame of a fixation.

The success of letter recognition depends on how text falls within the reader’s perceptual span, the effective visual area used during a fixation. Important terms like foveal vision¹ and parafoveal vision² (Figure 14) are extremely important concepts as they affect the users ability to identify letters.



3. Crowding is an interference effect where a character that is perfectly recognizable in isolation becomes illegible when closely surrounded by neighboring letters (Pelli et al., 2007). This is not caused by simple masking or low acuity, but by a mandatory process in which the brain inefficiently integrates features from the target and the flankers.

Fig.14. Cone of Vision showing foveal, parafoveal, near peripheral and peripheral vision. Solso (1994).

The most significant perceptual constraint challenging both the participants and VI-OCR, especially in the parafoveal field, is crowding. The typographic manifestation of crowding³ is particularly acute when the space between characters (tracking and kerning) is inadequate. If the inter-letter spacing is too tight (Figure 15), the individual letterforms merge into an indistinguishable visual texture, often referred to as a “barcode effect” (Beier, 2012).

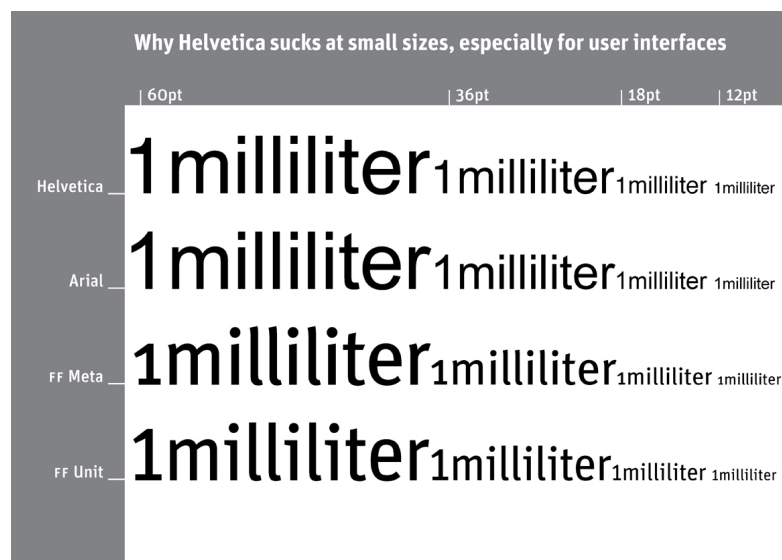


Fig.15. Why Helvetica sucks at small sizes, especially for user interfaces. Spiekermann (2012). <https://spiekermann.com/en/helvetica-sucks/>

Research has shown that even subtle adjustments to spacing can have a profound impact; studies by Tai, Sheedy, and Hayes (2006) confirmed that minor increases in letter spacing significantly facilitate word encoding and improve reading speed, especially for readers whose vision is already compromised. Therefore, the design of the negative space, the counters and the separation between characters, must be managed meticulously to mitigate the crowding effect, ensuring that the letterforms remain distinct and segmentable even in

the rapid, lower-resolution environment of the parafoveal span (Beier & Larson, 2010).

3.2. Impact of Visual Impairments on Typeface Legibility

The study of accessible typography is fundamentally a study of compensatory design. The impact of visual impairment (VI) on typeface legibility is not merely a reduction in visual performance but a profound functional breakdown of the biomechanical and cognitive systems required for fluent reading, as discussed in the preceding section. For this research, as previously established, the test group is derived from the globally defined population of individuals experiencing functional vision loss.

The WHO as seen in Fig.2. provides a studied framework for classifying visual function, focusing on the operational severity of impairment. The target population for accessible typography, falls within the spectrum of Mild, Moderate, and Severe Visual Impairment, collectively referred to as Low Vision.

The target study group occupies a VI spectrum where the retained vision is sufficient for reading, which allows us to conduct typeface legibility testing. The core issue for the visually impaired readers is that their physiological loss directly amplifies the spatial constraints of the reading eye. This degradation manifests through a loss of fine detail resolution and contrast. Testing with real participants also means that you have variables that you can't possibly control, either from the testing environment conditions (such as how sunny it was on that day in that specific room) or either from the participants themselves.

Regarding the participants individual conditions these vary according to a plethora of factors, although we don't have the medical expertise to fully understand these, from our interview questions with the participants, we gathered interesting conclusions with most of them reporting that their day-to-day vision varies (even if only a little bit) and consequently this affects their day to day lives. Their own impairments may become worse over the years, with institutions like ACAPO, providing VI people with tools to help them accept their own individual constraints whilst remaining as independent as possible. Although the target of this study remains the aforementioned group of people, most of these will eventually see their vision becoming worse over the passing of time, losing the ability to see the detail needed to identify text and other elements of their day to day lives.

Regarding the participants and the presented context of this study, it is important to understand a few key concepts. Reduced Visual Acuity means that the individual character must be presented at an enlarged size to meet the minimum resolution threshold of the impaired retina (Legge & Bigelow, 2011). In macular conditions, the damage to the central fovea forces the reader to utilize an eccentric retinal location for fixation, an area of the retina never intended for high-acuity reading. This eccentric viewing strategy immediately and drastically reduces the effective reading speed because the reader is unable to capture sufficient information within the parafoveal span to accurately program the next saccade. Contrast Sensitivity (CS) is the ability to perceive subtle differences in brightness. This is critical because all aspects of typography (stroke, aperture, and counterform) are defined by boundaries of contrast. In Low Vision, a reduced CS means that a typeface that appears sharp and dark to a normally sighted reader will appear washed out, ghosted, or blurred. This effect is particularly damaging to typefaces with thin strokes and low inter-stroke contrast.

Different eye pathologies degrade visual perception in distinct, yet typographically significant, ways. Understanding these mechanisms is essential for de-

signing resilient letterforms. Cataracts are the clouding of the lens, causing light scatter and reduced contrast. The scattering of light causes “veiling glare” that reduces overall contrast and makes counterforms appear to “fill in,” collapsing the aperture (Beier, 2012). Glaucoma is a progressive damage to the optic nerve, often leading to loss of peripheral vision (visual field loss). Loss of peripheral vision eliminates the parafoveal cues needed for accurate saccade programming, forcing the reader into a slower, word-by-word reading strategy (Rayner & Pollatsek, 1989). Macular Degeneration refers to the damage to the fovea (central retina), often creating a central scotoma (blind spot). Forces eccentric viewing, relying on peripheral vision that is inherently more susceptible to the crowding effect and has lower innate acuity.

A key consequence of conditions causing blur (like cataracts) or reduced acuity (like MD) is the collapse of negative space. The brain’s failure to sharply resolve the boundary between the stroke and the white counterform causes the stroke to spread perceptually. Typographically, this phenomenon collapses apertures (In letters like ‘c’, ‘e’, and ‘a’, the counter space (aperture) fills with perceived ink, making it impossible to differentiate these letters from closed forms like ‘o’ or ‘d’) and reduces distinctiveness (The visual ambiguity caused by this collapse transforms recognition tasks into error-prone guessing, leading to the frequent misrecognition of visually similar characters (Beier & Larson, 2010). For the Low Vision reader, the phenomenon of crowding, where letters are easily recognizable in isolation but illegible when closely flanked, is amplified to catastrophic levels. As Pelli et al. (2007) established, crowding is not a simple acuity issue but a failure of the brain to properly segment features in the visual periphery.

In Low Vision, especially with eccentric viewing, the critical spatial frequency of the visual system is expanded, meaning that the necessary minimum distance between the target letter and its neighboring flankers is significantly larger than for a normally sighted individual. If the inter-letter spacing is set for a normal reader, the impaired visual system cannot isolate the individual characters. This forces the typographic solution to become a spatial one.

It is also important to note that from the self-reporting of their visual conditions we also found out, during our interviews, that the visual conditions of the participants, although stable, are incredibly different between each individual, with most of them having a combination of several types of visual problems.

Therefore, for the Low Vision reader, the conventional typographic standards that govern line length, size, and spacing are rendered differently. The design must be optimized to overcome the physiological failures of the eye. Perhaps this optimization has to allow the individual to customize the typeface within certain limitations, transforming it from a subtle medium into an explicit tool for visual compensation.

3.3. Contemporary Research on Improving Legibility

The contemporary research landscape on typeface legibility has moved from generalized to empirical, evidence-based science. As previously mentioned these studies establish which typographic variables can be successfully optimized to enhance reading performance, particularly for the visually impaired (VI).

Historically, much of the research on legibility was dedicated to determining whether serif or sans-serif typefaces were inherently superior. Early, influential studies yielded often contradictory and inconclusive results. Roethlein’s early work (1912) suggested that certain serif faces were measurably more legible than their sans-serif counterparts, a finding often cited to support the historical preference for serif typefaces in body text and later on contradicted by authors

such as Tinker and Paterson (1932), in one of the most exhaustive early studies, found only negligible differences in reading speed between common serif and sans-serif faces, suggesting that factors other than style were more important.

This ongoing debate over style (serif vs. sans-serif) was largely rendered obsolete by the dominance of the letter recognition model in cognitive psychology (Larson, 2004; Rayner et al., 2012). Contemporary research established that the primary factor governing legibility is not the presence of a serif, but the distinctiveness, simplicity, and consistency of the individual letter's features. The scientific consensus now holds that a well-designed sans-serif face can be as, or more, legible than a poorly designed serif face, provided its critical features are optimized. The focus shifted from what the typeface is to how its components function under visual stress.

Contemporary legibility studies prioritize isolating and optimizing specific typographic parameters to compensate for the physiological losses inherent in visual impairment.

A critical focus is on reducing the confusion matrix among easily misrecognized letters, especially in the low-resolution environment of the parafoveal span or the compromised retina of a VI reader. Beier and Larson (2010) conducted seminal research, showing that the design of specific features directly impacts recognition accuracy.

The width and openness of the apertures (the openings in letters like 'c', 'e', 'a', 's') were proven to be vital. Wide apertures make these letters resistant to the blurring and "fill-in" effect common in cataracts or low-contrast viewing. The researchers designed specific improvements for high-frequency error letters to enhance their uniqueness. Maintaining a high x-height ensures maximum resolution under conditions of reduced VA (Legge & Bigelow, 2011). Simultaneously, the stroke weight must be robust to maintain contrast visibility (CS) without causing the counterforms to collapse. Optimized accessible typefaces must employ generous tracking (increased horizontal distance between all characters) to spatially separate the characters beyond the crowding zone. The letterforms themselves must have exaggerated differences, for instance, a wide-open 'c' and a clearly circular 'o', to provide maximum differentiation even when degraded by blur and low contrast (Larson, 2004).

The single most critical adjustment for improving legibility in the VI population is the manipulation of horizontal spacing. Perea and Gomez (2012), demonstrated that even subtle increases in inter-letter spacing (tracking) measurably facilitated the encoding of words in normally sighted individuals, confirming the spatial nature of word processing. For VI readers, this subtle benefit becomes a necessity. Studies concerning reading performance in patients with central vision loss (e.g., macular degeneration) consistently show that increased spacing (often far beyond conventional aesthetic standards) is the dominant factor that restores reading function. By providing extra space, the typeface spatially separates the letterforms, moving them outside the VI reader's expanded crowding zone and allowing for accurate character segmentation. The research concludes that for accessibility, the primary legibility variable is not the letter shape itself, but the white space surrounding the letter (Beier, 2012).

Perhaps the most significant development is the integration of empirical testing directly into the typeface design process. The Sitka Project, a collaboration between Larson and Carter (2016), stands as an example of this.

Conventional typefaces are often designed at one size (e.g., 12pt) and then mechanically scaled up or down, which often renders the face illegible at extreme sizes. The historical solution, Optical Scaling (Carter, 1984), involved designing different masters for different sizes. Larson and Carter applied this optical scaling principle not just for aesthetic reasons, but for scientific, functional performance. They created six different masters (Caption, Text, Subhead, etc.), each optimized for a specific reading size. They subjected these masters to iterative, scientifically rigorous legibility testing, adjusting features like aperture and stroke weight based on quantified error rates. The project successfully demonstrated that typeface design could be a scientific feedback loop, resulting in a family of faces where each master provided superior legibility within its target size range compared to simple mechanical scaling.

The future of legibility research involves standardizing metrics and employing technology to simulate and assess visual performance. The International Organization for Standardization (ISO) published the ISO 24509 (2019), which provides a normative method for estimating the minimum legible font size for people of any age, but not for people with visual impairments such as low vision. Although this type of framework moves legibility assessment from anecdotal testing to a formalized, internationally agreed-upon metric based on the functional relationship between visual acuity, luminance, and character size, it is important that an international standard takes into consideration VI individuals.

As mentioned before in previous chapters the VI-OCR (Visually Impaired Optical Character Recognition) pipeline (Gao et al., 2025) and other similar solutions use computer vision models to simulate the effects of specific visual impairments (like reduced CS and high spatial frequency loss) on text recognition. This, in theory, will allow rapid testing and quantifying of the resilience of a typeface's design features against various VI conditions, providing immediate data on how blur or reduced contrast compromises apertures and counterforms. These computational tools may represent the next phase in legibility science, offering a powerful method to predict typographic performance under stress.

3.4. Typeface Legibility Assessment Methods

The assessment of typeface legibility is a structured, multi-stage process that transitions from the qualitative evaluation of design integrity to the quantitative measurement of user performance. A rigorous legibility study requires both initial design validation (Typeface Quality Assessment) and subsequent empirical testing (Legibility Testing) to ensure that observed performance differences are attributable to the tested variable and not inherent design flaws.

The quality of a typeface must be established before its legibility can be tested, a poorly constructed font may introduce artifacts or inconsistencies that skew empirical results. Typeface quality, as distinct from legibility, refers to the systematic rigor, geometric precision, and structural resilience of the design.

Type designer and researcher Gerry Leonidas, who agreed to answer our questions, provides a critical framework for this initial assessment, emphasizing that quality resides in the system and the consistency of the execution (Leonidas, 2013).

Quality demands that the relationships between characters are consistent and mathematically sound. If a design has multiple weights (Light, Regular, Bold), the transition between these masters must be flawless. Any inconsistency in the interpolation (the automated creation of intermediate weights) can introduce structural deformities, especially in apertures or counterforms, which are det-

perimental to legibility for the visually impaired (VI). In his research, Leonidas stresses that the meticulous definition of sidebearings (the space on either side of a character) and kerning (the space between specific character pairs) is foundational (Leonidas, 2015). If the initial spacing is erratic, the visual rhythm of the text breaks down, increasing fixation variability and reducing reading efficiency. The design must be robust enough to survive low-resolution rendering, sub-pixel rendering effects, and the blurring caused by vision loss.

Once the structural quality is verified, functional legibility is quantified using objective, performance-based metrics derived from psychophysics and cognitive psychology. The most common method of legibility assessment involves measuring how quickly and accurately a reader processes a given text, typically expressed in Words Per Minute (WPM). Historically, studies used standard blocks of text, measuring total reading time and comprehension errors. The early work by Tinker and Paterson (1932) defined this methodology but often failed to isolate variables effectively, leading to the conclusion that font style differences were negligible.

Later, authors such as Legge & Bigelow (2011), reference specialized reading charts like, for example, the MNREAD Acuity Chart. This chart presents short sentences in high-contrast text at systematically decreasing sizes, measuring three crucial metrics - Critical Print Size (CPS), Maximum Reading Speed (MRS) and Acuity Reserve. MNREAD and modern standards rely on LogMAR (Logarithm of the Minimum Angle of Resolution) scaling, which ensures that each step in print size represents a proportional loss of resolution, allowing for a standardized, linear comparison of acuity across different reader populations.

Kevin Larson's proposed methodology involves the presentation of single letters or short non-words (e.g., flashing the stimulus for 50ms). By forcing the reader to rely on a brief, unassisted fixation, this technique generates a misrecognition matrix, revealing which characters are most frequently confused (e.g., 'c' for 'o', 'h' for 'n'). This empirical data allows designers, as demonstrated in the research by Beier & Larson (2010), to target and redesign the most error-prone characters.

Legibility studies must assess a typeface's performance under crowding (Pelli et al., 2007). This involves presenting a target letter flanked by other letters and systematically increasing the spacing until the target letter is recognizable. For the VI population, the critical spacing (the distance needed to eliminate crowding) is measured, providing direct evidence on how much horizontal space the typeface requires to be functional.

Despite the rise of objective metrics, a necessary critical discussion persists regarding the scientific validity of these tests and the role of reader preference. Critics argue that metrics like WPM and misrecognition rates, while objective, fail to capture the holistic reading experience. A typeface optimized purely for speed (e.g., through extreme magnification and spacing) may feel alienating or cumbersome in a natural reading environment (Unger, 2007). Early research, such as Roethlein's (1912), often favored traditional serif typefaces based on self-report and preference, even when objective speed differences were minor.

While modern science largely discounts preference as the primary driver of legibility, the most effective contemporary protocols of assessment of typeface legibility propose to combine these methodologies, exemplified by the development of the Sitka typeface. The Sitka project demonstrated an iterative cycle where design adjustments based on misrecognition data led to the creation of optically scaled masters. Each master was then validated through WPM and acuity testing, successfully integrating empirical science into the design process.

Typeface legibility simulation models that are emerging, such as the previously mentioned proposed VI-OCR pipeline (Gao et al., 2025), offer a predictive, non-human method for initial assessment. By using AI models to simulate the visual effects of VI pathologies and quantifying the resulting character recognition errors, there's potential to quickly and reliably forecast a typeface's resilience without the expense and time of large-scale human trials. The key missing factor in these types of pipelines is to prove that they can accurately emulate the same type of results that can be obtained with human test subjects. If the "machine" fails a lot or is very accurate, never failing, then it cannot simulate human vision with precision. The goal, therefore, should be to create a pipeline that fails as much as we do, as humans, in order to create an accurate, precise, and valid emulation of visual conditions.

The assessment of typeface legibility requires a multi-faceted methodology that comprises the validation of the design's structural integrity. But it shouldn't end here. Computational simulations allow for the rapid and low-cost quantification of a typeface's resilience under stress, provided they can accurately emulate human recognition errors rather than achieving perfect machine recognition. This multifaceted approach ensures that the final design is grounded in empirical evidence and functional reliability.

If a proposed simulation pipeline is tested and proven we should integrate it to ensure that the typeface is resilient to the challenges of VI users. Although simulation models are interesting and have an incredible amount of potential we should never forget that human testing should be run alongside it (using psychophysical testing), providing the designer with comparative quantitative information, which then can be applied to the design of a typeface.

4.

Computational Tools and Models for Legibility

Chapter Summary:

→ This chapter examines the shift from subjective legibility assessment toward evidence-based computational modeling, bridging the gap between physiological optics and cognitive recognition. The foundation of this approach lies in creating scientifically grounded approximations of the low-vision retinal image through two primary mathematical frameworks: the Contrast Sensitivity Function (CSF) and the Point Spread Function (PSF). By utilizing Fourier analysis, the CSF models the eye as a low-pass filter that attenuates the high-frequency components necessary for resolving fine typographic details, while the PSF simulates the complex light scatter and “veiling glare” characteristic of conditions like cataracts. These models demonstrate the physical “collapse” of counterforms and apertures, providing objective evidence for design strategies that prioritize character distinctiveness and generous spacing.

The research further integrates neural and spatial constraints, moving beyond uniform blur to account for the non-uniformity of the human visual field. This involves modeling eccentricity-dependent degradation, where visual resolution drops sharply outside the central fovea, and the introduction of Equivalent Input Noise (EIN) to represent the stochastic “noise” of a compromised neural pathway. A critical focus is placed on the phenomenon of visual crowding (the failure to isolate adjacent features in the periphery). Because simple image filters cannot replicate this cognitive failure, the chapter argues for a combined approach that couples physiological simulation with a pattern recognition surrogate.

The final section positions modern Optical Character Recognition (OCR) as this necessary cognitive proxy. Tracing the technological evolution from brittle, segmentation-dependent classical systems to robust, end-to-end Deep Learning and Transformer-based architectures, the text explains how OCR performance serves as a quantifiable metric for human legibility. Because modern engines like TrOCR can achieve near-human performance on ideal text, their failure patterns on degraded images, measured through Character Error Rates (CER), directly mimic human recognition errors. This integrated pipeline allows for the high-throughput, objective assessment of typographic resilience, transforming computational simulation into a predictive tool for both accessible design optimization and the creation of immersive empathy-driven interfaces.

4.1. Foundational Models of Visual Degradation Simulation

The foundation of computational legibility assessment lies in the scientifically robust method for replicating the image quality experienced by the low-acuity reader. This simulation can't be a simple digital filter, it needs to be a complex, multi-layered physiological and optical modeling process. The ultimate goal is to produce a simulated retinal image that accurately reflects the frequency, contrast, and spatial degradation caused by specific ocular pathologies, thereby generating the necessary input for a cognitive recognition engine (Gao et al., 2025).

The Contrast Sensitivity Function (CSF)

The Contrast Sensitivity Function (CSF) stands as the most established and widely used psychophysical model for simulating the effects of degraded vision (Pelli & Bex, 2013; Legge & Bigelow, 2011). Unlike basic Visual Acuity (VA) measurements (e.g., the Snellen fraction), which only quantify the ability to resolve the smallest detail at maximum contrast, the CSF provides a comprehensive measurement of the visual system's sensitivity across the entire spectrum of spatial frequencies (Pelli & Bex, 2013). This spectrum is crucial because a typeface, or any image, is mathematically composed of a variety of frequencies (Pelli & Bex, 2013), high frequencies convey sharp edges, fine serifs, and thin strokes, while low frequencies define the overall structure, x-height, and broad letterform contours.

The implementation of the CSF model relies on linear systems theory and Fourier analysis (Pelli & Bex, 2013; Campbell & Robson, 1968), requiring a baseline understanding of how visual information is processed within the frequency domain (Figure 16). Specifically, Fourier decomposition allows any complex visual input to be transformed into the frequency domain, where it is represented as a sum of simple sinusoidal waves of varying spatial frequencies, amplitudes, and orientations (Pelli & Bex, 2013).

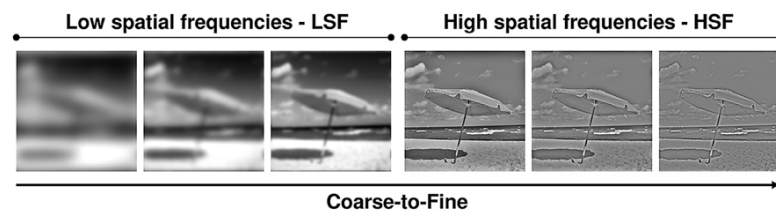


Fig.16. Coarse-to-fine sequence of spatial frequency processing (from low-to-high spatial frequencies) during scene perception. Kauffmann et al. (2014).

The neural and optical components of the healthy human eye act as a bandpass filter that allows mid-range frequencies, those essential for rapid form recognition, to pass through most efficiently (Pelli & Bex, 2013). However, ocular pathologies such as cataracts, optic neuropathy, or high refractive errors (Figure 17) depress and narrow this natural bandpass filter by functioning as powerful low-pass filters that drastically attenuate high spatial frequencies (Legge & Bigelow, 2011). This mathematical process effectively simulates the loss of detail inherent to the condition (Pelli & Bex, 2013).



Fig. 17. Eye Disease Simulations. Images from the Public Domain, created by the National Eye Institute, United States Government.

A specific degradation filter, derived from a patient's clinically measured or simulated Contrast Sensitivity Function (CSF) curve, such as a known curve for 20/100 vision, is applied to the Fourier components of the text image (Pelli & Bex, 2013). This process modulates the contrast and amplitude of specific frequency components relative to the human data (Legge & Bigelow, 2011). When the resulting image is reconstructed via the inverse Fourier transform, it yields a blurred, low-contrast output that serves as a scientifically grounded approximation of the retinal image perceived by a patient with that specific impairment (Gao et al., 2025). This method is the backbone of models used in many industrial and ergonomic standards, such as ISO 24509:2019, which uses VA and luminance factors (which rely implicitly on CSF-related principles) to estimate minimum legible font size (ISO 24509, 2019).

The Point Spread Function (PSF)

While the CSF model is highly effective for simulating uniform blur and frequency loss, it is insufficient for modeling the complex, light-dependent phenomena of glare and light scatter (Pelli & Bex, 2013). These are critical factors in conditions affecting the media of the eye, such as advanced cataracts, corneal edema, or posterior capsule opacification (haze) (Legge & Bigelow, 2011). These simulations require the application of the Point Spread Function (PSF) (Legge & Bigelow, 2011).

The PSF describes how the optical system distributes light originating from a theoretical point source (Pelli & Bex, 2013). In a pristine, healthy eye, the PSF is a tight, small function. In an impaired eye, light is scattered away from its original point, creating halos, starbursts, and a general reduction in boundary contrast (Figure 18) (Legge & Bigelow, 2011; Gao et al., 2025).

When high-contrast text (e.g., black text on a bright screen) is viewed, the bright background produces scattered light that spills into the dark strokes of the letters. This phenomenon, known as veiling luminance or disability glare, effectively reduces the effective contrast of the text boundary (Legge & Bigelow, 2011). The simulation is performed by computationally convolving the text image with a broad, scatter-driven PSF kernel (Pelli & Bex, 2013; Gao et al., 2025). The shape of this kernel is often empirically derived from clinical measurements of light scatter in various patient populations to ensure pathological fidelity (Legge & Bigelow, 2011). The integration of the PSF model provides vital, actionable feedback for type designers that CSF models often overlook (Beier & Larson, 2010; Larson & Carter, 2016). It clearly demonstrates the computational collapse of crucial typographic features, such as

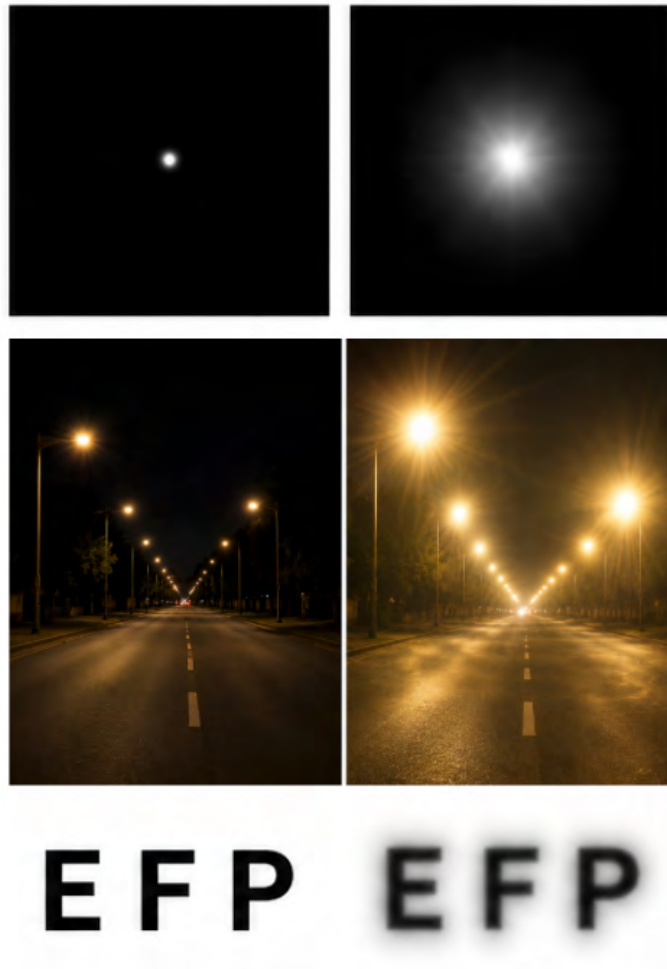


Fig.18. Visual effects of PSF.

counterforms and apertures, due to light spillover (Legge & Bigelow, 2011). This provides objective evidence for the necessity of design strategies focused on high distinctiveness (Beier & Larson, 2010) and generous spacing to resist glare and maintain feature isolation (Larson & Carter, 2016).

Neural System Modeling: Eccentricity and Crowding

To move beyond uniform optical degradation, foundational models have also sought to replicate the spatial and temporal limits of the neural visual system (Pelli & Bex, 2013). The primary mechanism for this replication is by modeling the non-uniformity of the visual field, that is, how visual performance degrades systematically (Figure 19) as a function of eccentricity (distance from the center of gaze) (Legge & Bigelow, 2011). This addresses the post-retinal limitations of feature recognition, such as visual crowding, the inability to identify objects in visual clutter, which sets a fundamental limit on conscious perception throughout the periphery (Whitney & Levi, 2011).

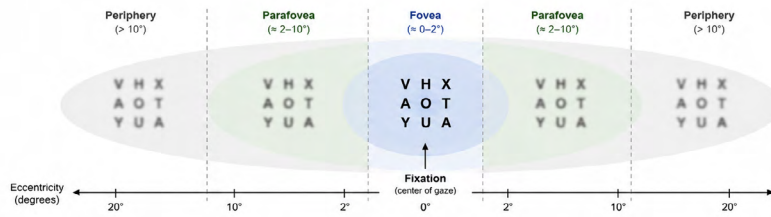


Fig.19. *Effect of Eccentricity on Letter Recognition Across the Visual Field.*

Non-Uniform Visual Space and Eccentricity Scaling

A core physiological limitation of human vision is the non-uniform distribution of photoreceptors and neural processing units across the retina. The human retina clusters the majority of its high-resolution cone photoreceptors in the fovea (central 1° to 2° of vision) (Pelli & Bex, 2013). Beyond this point, receptor density plummets, and the neural integration zones (receptive fields) become much larger (Legge & Bigelow, 2011). Computational models simulate this by applying an eccentricity-dependent degradation kernel (Pelli et al., 2007; Legge & Bigelow, 2011). Text or visual stimuli positioned centrally receive a minimal degradation filter, while text presented in the simulated periphery is subjected to progressively massive degradation (blur, contrast loss, and increased pooling) (Pelli et al., 2007; Legge & Bigelow, 2011). This model, which ensures that visual performance is systematically worse outside of the central fixation point, provides a crucial layer of perceptual fidelity for typographic assessment (Larson & Carter, 2016). This accurately reflects the functional fact that peripheral visual acuity drops off rapidly (Pelli et al., 2007).

By integrating eccentricity into the model, researchers can accurately predict the Critical Print Size (CPS), the smallest point size at which reading rate reaches its maximum, and the Critical Spacing required to maintain character segmentation (Legge & Bigelow, 2011), reinforcing the argument for generous letter and word spacing in accessible design (Pelli & Tillman, 2008).

Equivalent Input Noise (EIN) Models

To simulate the functional, non-optical inefficiency of the neural system, some predictive models incorporate the concept of Equivalent Input Noise (EIN) (Pelli et al., 2007). This is a computational technique used to model the cumulative effect of neural processing failures, the random “noise” that the brain must filter through to extract meaningful information (Pelli et al., 2007).

EIN models introduce a controlled level of random visual noise into the simulated retinal image (Pelli et al., 2007). This noise is designed not to mimic external static but rather the internal stochasticity and decreased signal-to-noise ratio inherent to a compromised visual pathway (Pelli et al., 2007).

By calibrating the level of EIN, researchers can create a model that struggles to read text even when the optical image is moderately clear (Pelli et al., 2007). This technique is valuable for predicting the functional performance loss that exceeds simple optical calculation (e.g., in early-stage glaucoma or central field loss), where the brain receives fragmented or corrupted information despite an optically decent image (Pelli et al., 2007).

The Cognitive Barrier: Visual Crowding and Recognition

Despite the precision of the CSF, PSF, and retinal sampling models, they share a profound, fundamental limitation: they model only the retinal input and entirely fail to simulate the subsequent, essential neural processing failures, the “cognitive” part of reading (Larson & Carter, 2016).

The most challenging phenomenon for typography simulation is visual crowding (Beier et al., 2022; Xiong et al., 2021). Crowding is not caused by blur, it is a breakdown of the neural mechanism responsible for segmenting and isolating features (Figure 20) (Pelli et al., 2004; Whitney & Levi, 2011). It is a mandatory integration of visual elements in the periphery, resulting in the failure to accurately recognize a target letter when flanked by similar adjacent letters (Pelli et al., 2007).

r + are

Fig.20. Demonstrating and mapping the spatial limits of visual crowding. Pelli et al. (2007)

As established by the research (Pelli & Bex, 2013), no amount of simple CSF-based blurring can accurately reproduce the specific, catastrophic interference pattern of crowding. A blurred 'e' may look ambiguous, but it does not merge into an unreadable “mongrel texture” with the adjacent letters as crowding demands (Whitney & Levi, 2011).

Because reading is an active cognitive process involving fixation, attention, and recognition (Rayner et al., 2012), any effective computational tool must move beyond merely simulating the degraded image. It must couple the precise physiological simulation (CSF/PSF) with a cognitive surrogate, a pattern recognition unit capable of exhibiting human-like confusion and error patterns (Gao et al., 2025; Larson, 2004). This profound limitation of purely physiological models, their inability to simulate the neural failure of crowding and subsequent recognition, is the primary motivator for the development of integrated, machine learning-based approaches (Gao et al., 2025).

4.2. Precursors and Alternative Computational Legibility Assessment Approaches

The challenge of computational legibility assessment requires overcoming a fundamental limitation - purely optical and physiological models (like the Contrast Sensitivity Function, CSF, and Point Spread Function, PSF, discussed in Section 1.12) effectively simulate the physical degradation of the retinal image but fail to account for post-retinal, or neural, limitations (Larson, 2004). Among these neural failures, as previously mentioned is visual crowding, the failure to recognize objects in visual clutter, which sets a mandatory limit on conscious perception in the periphery (Whitney & Levi, 2011). This type of disconnection means that the simulation models must include a cognitive surrogate capable of predicting recognition failures.

The evolution of computational legibility pipelines was built on decades of psychophysical research that established quantitative, predictable limits on human reading performance. With this in mind, any successful model must accurately predict several metrics. Vision science established that the reading rate remains constant above a specific threshold print size known as the Critical Print Size (CPS), and plateaus at a fixed Maximum Reading Rate (MRR) (Legge & Bigelow, 2011; Pelli et al., 2007). The ability to predict these human-measured thresholds became the standard for validating computational model accuracy. Studies in the psychology of reading provided precise spatial parameters for the perceptual span, the window of text processed during a single fixation, which is drastically reduced in low vision (Rayner et al., 2012).

This research reinforced the need for computational models to simulate the rapid drop-off in acuity and sensitivity outside the central fovea (Pelli et al., 2007).

The earliest simulations relied on applying uniform blur, but the realization that pathology-specific, spatially variant filtering was required led to the widespread adoption of the CSF-based filter (Pelli & Bex, 2013). The CSF filter operates by mathematically translating clinical measurements (Visual Acuity and Contrast Sensitivity) into parameters for a frequency-domain filter. As described by Thompson et al. (2017), the simulation is parameterized by taking standard clinical measurements (e.g., 20/200 visual acuity) and using them to define the magnitude and position of the CSF curve shift. A reduction in acuity corresponds to a horizontal shift (attenuating high spatial frequencies), and a loss of contrast sensitivity corresponds to a vertical depression of the curve. This mathematical process models the reduced spatial-frequency visibility inherent to the impairment, thereby predicting how a low-vision observer perceives text and environments (Thompson et al., 2017). The predictive utility of these CSF filters was established by rigorous validation. Xiong et al. (2021) implemented a low-vision image filter based on clinical CSF data and performed a critical experiment: they compared normally sighted subjects viewing text degraded by the filter to real low-vision readers. The results showed strong agreement, with the simulated low-vision readers exhibiting reading acuity and critical print sizes very similar to actual low-vision subjects up to 2.0 logMAR (Xiong et al., 2021). This success confirmed that the optical filtering stage is a valid proxy for retinal input loss.

However, since this filter only models the retinal input, it remains unable to predict neural failure modes like visual crowding or the subsequent cognitive recognition process (Larson, 2004). The RetModel (Retinal Model), pioneered by Horváth and Dömötör (2017), is one of the most detailed precursors to modern pipelines. It was introduced as simulation software specifically designed for reproducible font legibility measurement. The model's key innovation lies in its sophisticated, non-uniform simulation of the retina. The RetModel goes beyond simple uniform blur by creating a high-resolution, non-uniform representation of the retina. This includes simulating the precise arrangement of photoreceptors and the pooling of neural signals across the visual field (Dömötör & Horváth, 2025). The model specifically simulates direction-dependent retinal acuity loss and eccentricity-dependent degradation, applying localized filtering based on a feature's simulated distance from the central fovea (Dömötör & Horváth, 2025). The resultant degraded image, which attempts to accurately represent the input received by the visual cortex, is then fed into an Optical Character Recognition (OCR) engine (Horváth & Dömötör, 2017). The OCR acts as the cognitive surrogate. The failure of the OCR engine to correctly identify letters in the degraded image serves as a direct, quantifiable prediction of human legibility failure.

The extended RetModel, applied to 22 different fonts, produced legibility rankings that showed strong consistency with human reading studies (Dömötör & Horváth, 2025). A parallel field of simulation focuses less on predicting reading rate and more on creating high-fidelity, real-time, and immersive experiences of vision loss, primarily for empathy and interface evaluation. These tools often support gaze-contingent effects and specific pathological symptoms:

Velázquez et al. (2016) described a real-time video-processing system, often used with a VR headset, for simulating various eye diseases such as macular degeneration, retinopathy, and glaucoma. This system applies image masks and real-time degradation via the Hadamard product (element-wise matrix multiplication) to simulate pathological effects.

A key benefit is allowing sighted users to “experience” common impairments, aiding in the design of accessible interfaces.

Modern simulators leverage eye-tracking and AR/VR technologies to create highly personalized, gaze-contingent experiences. Jones and Ometto (2018) presented a VR/AR system designed as an empathy and design tool to evaluate how impairments affect readability and interface use, supporting common defects like disability glare, field loss, and blur. An advanced example is Krösl et al.'s (2020) “CatARact,” an augmented reality system that models the complex, light-dependent nature of cataracts, simulating symptoms like blur, glare, and tint based on the user’s instantaneous gaze position. This level of realism allows for the evaluation of text and signage in dynamic environments.

The Visually Impaired Optical Character Recognition pipeline (VI-OCR) represents, in theory, a modern, high-performance evolution of these integrated methods (Gao et al., 2025). It combines clinically validated low-vision filtering with advanced machine learning-based OCR.

The proposed VI-OCR pipeline (Gao et al., 2025) formalizes a process that is necessary for automated legibility assessment. This proposal can be better understood in the next sub-chapter.

By leveraging machine learning as a robust proxy for the human pattern recognition system, pipelines like RetModel and VI-OCR represent the current state-of-the-art in creating validated, high-throughput computational tools for the precise assessment and optimization of typographic design (Larson, 2004).

4.3. Optical Character Recognition

Optical Character Recognition (OCR) is the electronic or mechanical conversion of images of typed, handwritten, or printed text into machine-encoded text. Historically, OCR was developed primarily for document digitization and data entry (Wang et al., 2021). However, in the context of computational legibility assessment, OCR takes on a different specialized role, potentially serving as the cognitive surrogate for the human pattern recognition system. By providing a measurable, quantifiable output, specifically the character error rate, on a degraded image, OCR bridges the gap between the physiological simulation of vision loss and the neural process of text recognition (Larson, 2004).

The Evolution of OCR Architectures

Early OCR systems relied on a multi-stage, classical pipeline approach. These systems achieved high accuracy (often exceeding 99%) on clean, high-quality printed text but proved brittle when faced with noise, perspective distortion, or unfamiliar layouts (Ye & Doermann, 2015). The traditional process was modular and sequential, relying heavily on accurate segmentation. The input image was first cleaned via deskewing and noise removal, then converted to a binary image. Layout analysis and segmentation then identified text regions, breaking the page into columns, paragraphs, lines, and finally, individual characters. This reliance on robust segmentation was the critical point of failure for noisy or non-standard documents (Ye & Doermann, 2015). Statistical learning methods, such as k-Nearest Neighbors (k-NN), Support Vector Machines (SVMs), or Hidden Markov Models (HMMs), were used for final classification based on structural features.

The open-source engine, Tesseract OCR, exemplifies the transition from classical to modern methodologies (Smith, 2007). In its earlier iterations, Tesseract’s architecture used connected-component analysis to find characters and an adaptive two-pass character classifier to refine results.

However, with the release of version 4.0 and subsequent updates, Tesseract shifted to a hybrid model, integrating a Long Short-Term Memory (LSTM) neural network. This update significantly improved its ability to recognize sequences of characters rather than isolated glyphs, though its performance remains highly dependent on the quality of the initial layout analysis (Smith, 2007).

Modern Deep Learning and Transformer Pipelines

Modern OCR is dominated by end-to-end deep neural networks that have largely eliminated the brittle, error-prone segmentation steps of classical pipelines. A pivotal advance was the combination of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Introduced by Shi et al. (2017), the Convolutional Recurrent Neural Network (CRNN) processes the entire word image as a unified sequence. In this architecture, CNNs act as feature extractors, while recurrent layers model the character sequence and contextual dependencies. By utilizing the Connectionist Temporal Classification (CTC) transcription layer, these models can be trained end-to-end without requiring explicit character-level segmentation, which significantly improves robustness in complex scenarios like scene text recognition (Shi et al., 2017; Wang et al., 2021).

The current state-of-the-art leverages the architecture popularized by the Transformer model, enabling superior context capture and parallel processing. Li et al. (2022) proposed TrOCR (Transformer OCR), an end-to-end model that uses a Vision Transformer (ViT) to encode the image and a separate text Transformer for decoding. This architecture excels at capturing global context and modeling long-range dependencies, overcoming the limitations of RNNs in processing irregular layouts (Afkari-Fahandari et al., 2025; Li et al., 2022). These models often match or exceed human recognition rates on both printed and handwritten tasks.

Metrics, Benchmarks, and the Proxy Role

OCR performance is conventionally measured using error rates that quantify the distance between the output and the ground truth. The Character Error Rate (CER) is the ratio of character substitutions, deletions, and insertions required to transform the OCR output into the target text, divided by the total number of characters. Similarly, the Word Error Rate (WER) applies this logic at the word level. Researchers rely on standardized benchmarks, such as the ICDAR Robust Reading competitions and synthetic datasets like SynthText, to evaluate these algorithms. However, as noted by Neudecker et al. (2021), small failures in upstream layout analysis can disproportionately degrade these metrics, making consistent pre-processing essential.

Beyond document archiving and data entry, OCR plays a foundational role in assistive technologies. By enabling mobile apps to convert printed text into speech or Braille in real-time, OCR directly aids visually impaired individuals in navigating daily environments (Wang et al., 2021). In the scientific domain, the OCR engine's function is elevated to act as a pattern recognition proxy. Because an OCR model is trained to reach near-human performance on ideal text, its failures on mathematically degraded images—such as misclassifying “c” as “o” or merging “rn” into “m”—directly mimic human recognition failures caused by phenomena like visual crowding (Larson, 2004). This allows for the objective, scalable assessment of typographic resilience under simulated impairment.

5.

VI-OCR and its Application in Type Design

Chapter Summary:

→ The Visually Impaired Optical Character Recognition (VI-OCR) framework marks a significant shift in computer vision, transitioning from a “performance-first” approach to a “simulation-first” model. Instead of optimizing image clarity to enhance text recognition, VI-OCR functions as a modular, end-to-end pipeline that applies clinically-parameterized low-vision filters to typographic stimuli before passing them to a recognition engine. By processing images through a Fast Fourier Transform (FFT) to systematically remove high-frequency spatial information and depress the Contrast Sensitivity Function (CSF), the framework mathematically simulates the physical bottlenecks of human vision loss caused by conditions like cataracts or glaucoma. This workflow effectively treats the AI engine as a digital proxy for a human participant in a psychophysical experiment, allowing researchers to study exactly how visual impairments disrupt the reading process.

To achieve human-like error patterns, the pipeline utilizes state-of-the-art Transformer-based OCR (TrOCR) architectures. Because these models are trained on vast text datasets, they develop structural and cognitive archetypes that mirror human recognition, meaning they fail logically rather than randomly when faced with degraded images. This allows for the generation of Confusion Matrices that pinpoint critical character misidentifications (e.g., 'I' vs 'L' or 'e' vs 'o'). While setting up the official pipeline and its web interface (evaluatevisibility.com) presented initial technical challenges, prompting the temporary development of a proprietary, lightweight JavaScript simulator for rapid prototyping, transitioning to the original Python-based repository was necessary to ensure total methodological alignment and validate human-like failure patterns against actual human pre-tests.

Ultimately, VI-OCR alters the economics and speed of accessible type design by replacing intuition-led choices with an objective, virtual feedback loop during the active drawing phase. Instead of treating accessibility as an afterthought, designers can run micro-optimizations on letter anatomy or pair the pipeline with Variable Font technology to automatically scan parametric axes and find the optimal combination of weight, width, and optical size for legibility. While user interviews revealed some industry skepticism regarding the tool's immediate necessity, ethical validity, and practical integration, visually impaired individuals strongly welcomed the technology. The pipeline holds immense potential for creating pathology-specific typefaces tailored to the distinct visual noise of different ocular conditions, ensuring future typefaces can dynamically adapt to specific human needs.

5.1. The VI-OCR Framework: Architecture and Logic

The Visually Impaired Optical Character Recognition (VI-OCR) pipeline represents a specialized evolution of computer vision, designed specifically as an assessment tool for text accessibility (Gao et al., 2025). Unlike traditional OCR, which seeks to maximize accuracy through image enhancement, denoising, and binarization, the VI-OCR framework prioritizes “pathological fidelity.” It operates as a modular, end-to-end system where the input, in this case in the form of typographic stimuli, is subjected to a clinically-parameterized low-vision filter before being processed by a modern recognition engine. This fundamental shift from “performance-first” to “simulation-first” computer vision allows researchers to treat the recognition engine as a digital participant in a psychophysical experiment. By shifting the objective from “how can we read this best?” to “how does the impairment break the reading process?”, the framework provides a mirror to the human experience of low vision.

Throughout this document VI-OCR is often described as a “pipeline” or a “framework”, but what exactly does this mean? VI-OCR isn’t a tool, in truth it is, as continuously described, composed of multiple tools that constitute a workflow made of several different parts. If you want to “dumb this down” and make it a lot more visually easier to understand, VI-OCR can be broken down into two core parts. In Part I (Figure 21) we have an original image at a set resolution and size (in pixels), this image can be any type of image as long as it contains some type of text, in our test protocol the used image contained a trigram, in a specific typeface at a specific size (in points) in our case 120-24pt at 1920x1080 pixels and exported at 94PPI (to match monitor specifications). The original image is then filtered and consequently we get the output result. The filtration works according to several mathematical equations, as previously stated, this isn’t just a “blur effect”, it takes into consideration HShift, VShift, Distance from participant to the monitor (in our case 150cm), monitor size, in inches, (in our case 24in) and the screen size in pixels.

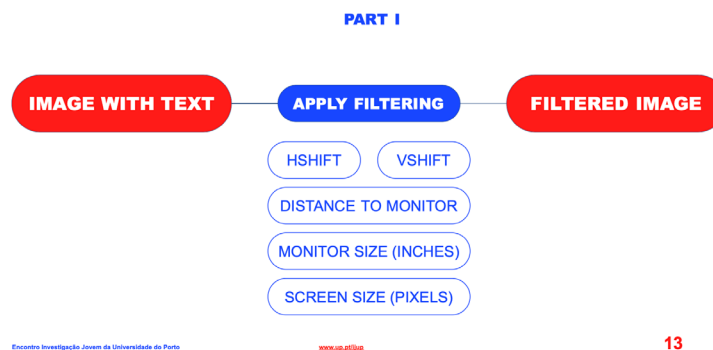


Fig.21. Slides from IJUP 2026, explaining the VI-OCR pipeline.

VI-OCR then applies the OCR to the filtered image to try and recognize text (Figure 22), emulating real participants, this being done through an AI model of our choice (in our case we conducted testing using Gemini). The output of this will be a folder, containing a text file, that presents the results in a specific format, like for example:

```
{
  "image": "10_96pt.jpg",
  "trigram": "['jXh']",
  "filtered_image_path": "outputs/filtered_images_Sub301/10_96pt.jpg",
  "hshift": 4.1684,
  "vshift": 0.0575
},
```

In this case at 96pt for image 10, with an user profile containing an HShift of 4.1684 and a VShift of 0.0575, VI-OCR recognized the trigram “jXh”.

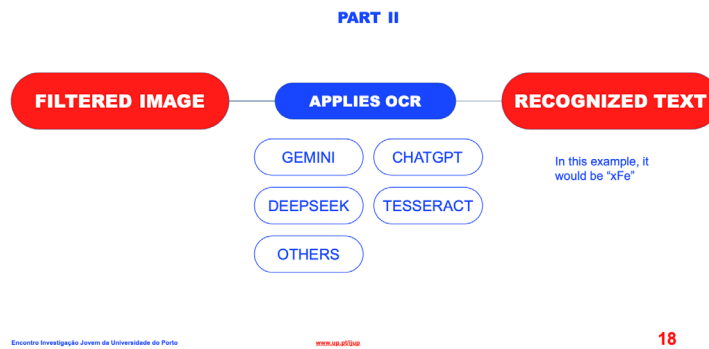


Fig. 22. Slides from IJUP 2026, explaining the VI-OCR pipeline.

Going into more detail, the technical foundation of this pipeline, as documented in the official repository (Gao, 2025), leverages a combination of vision science and deep learning. This architecture is built upon the premise that machine learning models, when sufficiently complex and trained on representative data, can serve as a robust proxy for the human pattern-recognition system. The pipeline is designed to be “end-to-end,” meaning it can ingest raw images of text, ranging from synthetic, high-contrast digital renders to complex, scene-based photographs, and output a character-level recognition analysis that includes confidence scores and error classification. This mimics the biological visual pathway: the input image represents the external world; the clinical filter represents the ocular media and retinal limitations (the “eye”); and the OCR engine represents the primary visual cortex and higher-level cognitive processing (the “brain”). For researchers, the utility of this framework is accessed either through cloning the existing repository and installing its heavy dependencies (such as PyTorch and specialized transformer libraries) or through specialized web-based interfaces, such as the evaluatevisibility.com platform. This website, which I was given access to by Dr. Xiong and Qingying Gao, allows for the precise modulation of visual degradation, defined by Visual Acuity (VA) and Contrast Sensitivity (CS) levels (Part I of the pipeline, Figure 21). This creates the “degraded retinal image” that serves as the basis for recognition testing, effectively simulating the physiological “bottleneck” of the human eye before the information reaches the “cortex” (the OCR engine).

Although this will be further explained in future chapters we ran into a lot of issues setting up the repository, using evaluatevisibility.com and coming up with alternative solutions that would allow us to proceed with the study without activating our, previously described, contingency plan. In an initial stage, while awaiting a response to my contact with the VI-OCR team, my dissertation supervisor, Professor Pedro Amado (Ph.D.), assisted in building a proprietary interface (Figure 23) based on the research published by Gao et al. (2025).

This web-based tool utilized a JavaScript-based implementation of Gaussian blur and contrast filters to mimic the initial findings of the VI-OCR team. This allowed for rapid prototyping and preliminary data collection within a lightweight environment, essentially providing a “proof of concept” for the computational approach. While this tool produced results highly consistent with the preliminary data, as the project progressed and I gained official access to the original Python-based software repository and the evaluatevisibility.com platform, a transition was made to the original pipeline. This move was essential to ensure total methodological alignment and to leverage the full testing and validation of the code structure, including the specific weights of

the TrOCR models and the precise Fast Fourier Transform (FFT) implementations, used in the peer-reviewed results presented in the VI-OCR paper. As previously mentioned, this transition was not worry-free and other alternative implementations based on the web-based tool were explored.

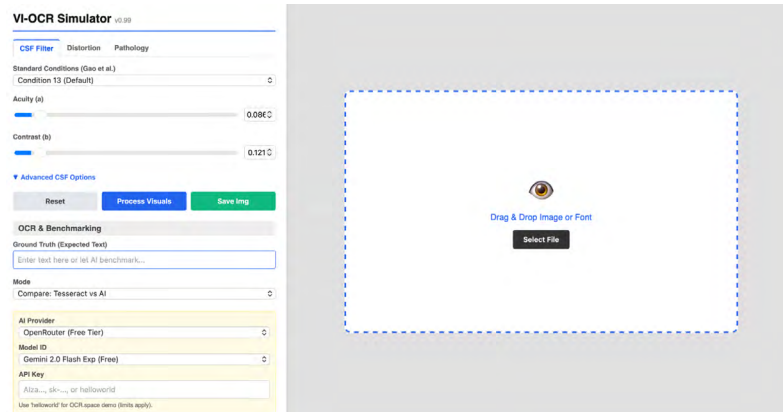


Fig.23. VI-OCR Simulator with Full Pipeline (Filter + OCR). Based on Gao et al. (2025). Source: Author. <https://typedesign.fba.up.pt/vi-ocr/>

5.2. Clinical Parameterization and Retinal Simulation

As previously mentioned, one of the core innovations of VI-OCR lies in its ability to translate human clinical data into digital image filters. This process, as mentioned before, moves from the “aesthetic blur” often used in design software, which usually relies on simple box or Gaussian blurs that do not account for the frequency-specific nature of vision loss. Instead, VI-OCR replaces these with filters derived from the Contrast Sensitivity Function (CSF) to simulate the experience of readers with various levels of impairment (WHO, 2019). This transition from arbitrary blur to parameterized simulation is what gives the pipeline its scientific validity and makes its results comparable to clinical vision trials.

The system applies a low-pass filter that systematically removes high-frequency spatial information. In the context of typeface design, high-frequency information is responsible for the “sharpness” and “definition” of character edges, thin serifs, and subtle transitions. When these are removed, the thin serifs of a high-contrast typeface (like a Bodoni) or the delicate terminals of a letter ‘a’ vanish, leaving only the “blobs” of the letter’s skeleton. The VI-OCR pipeline achieves this by converting the image into the frequency domain via a Fast Fourier Transform (FFT). In this domain, the image is represented as a collection of waves; the filter acts as a mask, multiplying the amplitude of high-frequency waves by zero or a near-zero fraction. By then performing an Inverse FFT, the system returns the image to the spatial domain, but now stripped of the details that a low-vision eye cannot resolve. This deconstruction allows the designer to see which parts of a letter are “visually expensive” and likely to fail first under stress.

By depressing the CSF curve, the model replicates the reduced sensitivity to light-dark transitions common in pathologies like glaucoma or diabetic retinopathy. This is not a simple “opacity” change where the white becomes grey, it is a frequency-specific reduction in amplitude. It simulates how, for a low-vision reader, a letter might be physically present on the page but lacks the edge contrast necessary to be “distinguished” from its background or neighboring characters. This results in a “ghosting” or “milky” effect where the text seems to wash out into the page, a phenomenon often described by patients with cataracts.

The evaluatevisibility.com tool, VI-OCR and our multiple test web-based solutions enables the researcher to set specific logMAR values (e.g., 0.4, 0.7, 1.0, 1.3) corresponding to a specific patient with a specific set of conditions. Unlike the Snellen scale (20/20), the logMAR scale provides a logarithmic progression that is mathematically ideal for statistical analysis. This scaling is usually utilized in vision-related tests. This precision allows for the calculation of a “Threshold logMAR”, the exact point where architectural integrity collapses under visual noise.

Although our own implementation of the test protocol will be explained in detail in future chapters, it is important to note that we aren’t utilizing logMar values, as our testing methodology is better suited for other types of analytical results. Also, although initially we intended to ask for help to conduct the assessment of the participants’ visual conditions we did not get a response from several people we contacted, opting instead for the participants to self report their visual conditions and to conduct a pre-test which will be detailed on future chapters.

5.3. OCR as a Proxy for Human Recognition Patterns

For the pipeline to be a valid simulation tool for human reading, the OCR engine must exhibit recognition errors that are qualitatively similar to human confusion. The VI-OCR pipeline utilizes state-of-the-art architectures, specifically the Transformer-based OCR (TrOCR) (Li et al., 2022). Unlike older OCR systems that relied on simple feature extraction or Convolutional Neural Networks (CNNs) that could be “tricked” by non-human patterns, TrOCR uses an “Image Transformer” for visual understanding and a “Text Transformer” for language modeling.

Because these models are trained on vast datasets of human-written and printed text, they develop a “learned” understanding of character structures that mirrors human cognitive archetypes. They do not just “see” pixels; they understand the structural relationships that define an ‘A’ or a ‘B’. When presented with a degraded image, these models do not fail randomly; they fail structurally and logically. For instance, if the bar of an ‘e’ is obliterated by simulated blur, the model, much like a human reader, may misidentify the character as an ‘o’ based on the remaining circular contour. This capacity for “human-like” error is what makes the VI-OCR a powerful tool for generating Confusion Matrices. These matrices provide a statistical map of legibility: they allow designers to identify “critical character pairs,” such as ‘a’ vs ‘o’, ‘i’ vs ‘l’, or ‘h’ vs ‘n’, that are most susceptible to legibility collapse. This empirical data provides a scientific justification for typographic adjustments, such as increasing apertures, exaggerating ascender heights, or differentiating the widths of similar characters, that intuition alone might miss. It effectively maps the “semantic distance” between letterforms and reveals where the “DNA” of one letter bleeds into another under visual stress.

In our own individual testing with four test typefaces, which will be detailed further on, we found that both VI-OCR and human participants often confused letters like ‘I’ and ‘L’ on specific trigrams. This is, of course, an expected result, but it shows the potential that VI-OCR has as a proxy for recognizing the same type of recognition and failure patterns as human participants. Although this is interesting, it is essential to note that VI-OCR in its current state is in active development, it can, and most likely will be continuously improved to make it better at identifying patterns and making the same type of mistakes under the same types of visual impairments.

5.4. The Utility of VI-OCR in the Type Design Process

If validated, the VI-OCR pipeline fundamentally alters the economics and speed of accessible type design. Traditionally, typeface designers have relied on “intuition-led” accessibility, applying general rules such as widening apertures or increasing x-heights without a granular way to measure their effectiveness during the drawing stage. This is something reinforced by the multiple interviews conducted in this study. Human testing, while the gold standard, can be expensive and slow, often happening only after a font is nearly finished, a point where significant architectural changes are difficult to implement. This often leads to “accessibility as an afterthought” rather than a foundational constraint.

VI-OCR bridges this gap by providing an immediate, objective feedback loop during the active design phase. The tool is particularly useful for comparative analysis. A designer can compare two variants of a character, for example, a ‘g’ with an open loop *versus* a closed loop, and see exactly in what conditions each version fails. This allows for “micro-optimizations” of letter anatomy that are invisible to the healthy eye but critical for the low-vision reader. This “virtual lab” environment means designers can iterate dozens of times in a single day, refining the “stress-resistance” of a letterform before ever involving human subjects. Furthermore, it allows for the testing of systemic legibility across a whole character set, identifying “weak links” in a font’s design that might cause recognition errors for specific pathologies. It transforms the drawing board into a testing ground where the designer can “hear” the machine struggle to read, much like a pilot uses a flight simulator to test a new aircraft’s limits.

Although, in theory, this is very interesting, in practice how could VI-OCR be applied and how does it fit the market and current landscape of the type design industry?

From the conducted interviews and more informal conversations in the context of this study, although designers often care about the legibility of their typefaces, they just develop them according to already established common knowledge, which isn’t exactly indicative their design is more or less legible than let’s say Helvetica or Times New Roman. Consequently, some designers are skeptical that a pipeline like VI-OCR can be useful to their own individual work and methodologies, or even ethically appropriate or “validated”. Whilst this is all very relevant information, it is important to note this is data from a fairly small number of conversations, it is impossible to understand the real impact of VI-OCR until it is “out there” in the market where you can hopefully get a whole different set of statistics.

Using Variable Font technology, a designer can generate hundreds of instances along axes such as Weight, Width, and Optical Size. VI-OCR can then act as an automated “tester,” scanning through the entire parametric space to find the “Legibility Sweet Spot”, the specific combination of weight and width that maintains the highest recognition rate under a certain threshold. This would replace the “guesswork” of choosing a default “Regular” weight for accessibility and allows for fonts that dynamically adapt to a user’s specific level of impairment. Another possible methodology could be based on pathology-specific optimization. Instead of designing a “generic” accessible font, designers could use the pipeline to optimize typefaces for specific conditions. By adjusting the CSF filter to mimic macular degeneration (loss of central vision) versus glaucoma (loss of peripheral vision), the tool could inform the creation of specialized typefaces optimized for the specific visual “noise” of those conditions. For example, a font for macular degeneration might prioritize high-contrast, large-scale peripheral features to aid eccentric viewing, while a font for cataracts might focus on hyper-generous spacing and “anti-glare” stroke weights to prevent light spill.

Regarding the VI individuals I talked with, which are also important in this context, all of them said that any tool that can help people with VI is welcome. None of them demonstrated any sort of worries regarding ethical questions that could potentially come out of the simulation of their vision through VI-OCR. They felt the research was important as it could lead to their day-to-day lives being improved. So, although we've theorized several possible use cases and methodologies, the important thing is that research for the visually impaired continues to evolve and that hopefully more typefaces are created and tested for them to use according to their type of visual impairment.

Literature Review Summary

The literature review establishes Type Design as a systematic discipline positioned at the intersection of visual communication, technology, and cognitive perception, emphasizing its fundamental objective of optimizing written communication for efficient and accurate reading. The review begins by defining the foundational concepts of typography, distinguishing between the abstract visual structure of the typeface and the functional digital font file that enables its implementation across contemporary media. Particular attention is given to the evolution of digital font technologies, including OpenType and Variable Fonts, whose flexible and programmable structures allow designers to manipulate typographic parameters with unprecedented precision. This technological evolution is presented as especially significant for contemporary legibility research, as it enables the controlled testing of incremental design variations across multiple visual conditions.

A central concern throughout this review is the distinction between legibility and readability. Legibility is framed as the perceptual clarity and recognizability of individual characters, while readability is understood as the broader reading experience involving text flow, comprehension, visual comfort, and reading efficiency. The literature review adopts an interdisciplinary perspective that understands readability as the result of the interaction between typographic design, layout structure, and the perceptual abilities of the reader. Within this framework, legibility becomes the foundational variable upon which successful reading performance depends, particularly in contexts involving visual impairment.

The literature review also situates typography within the broader context of inclusive and accessible design. It highlights how accessible typography seeks to compensate for the physiological and perceptual limitations associated with visual impairments such as cataracts, glaucoma, diabetic retinopathy, and macular degeneration. These conditions alter visual acuity, contrast sensitivity, peripheral vision, and the ability to distinguish fine typographic detail, thereby disrupting the biomechanical processes involved in fluent reading. The review explains how visual impairments intensify perceptual phenomena such as crowding, forcing designers to reconsider conventional typographic standards related to spacing, weight, and character differentiation.

Building upon this foundation, the review examines the biomechanics of reading and the cognitive processes involved in text recognition. Reading is described as a discontinuous process composed of rapid eye movements and fixation periods, during which visual information is acquired and processed. Contemporary cognitive research is presented in support of the letter recognition model, which argues that words are identified through the recognition of individual characters rather than overall word shapes. This perspective reinforces the importance of designing highly distinctive letterforms capable of maintaining clarity under degraded viewing conditions.

The review further explores the modern workflow of typeface development, presenting it as an iterative process involving conceptualization, drafting, production, validation, and distribution. In this context, type design is framed not only as an artistic practice but also as a form of digital systems engineering. Considerable attention is dedicated to the Validation phase, where typefaces are tested for technical stability, visual consistency, and functional legibility. Such testing is described as expensive, time-consuming, difficult to reproduce, and heavily influenced by participant variability and environmental factors.

It is within this context that computational simulation models emerge as a central focus of the review. Technologies such as VI-OCR and other vision-based computational simulation models are presented as promising tools capable

of simulating the perceptual effects of visual impairments on text recognition. These offer the potential to rapidly evaluate how typographic features behave under conditions of blur, reduced contrast sensitivity, and visual field loss. The literature positions these computational approaches not as replacements for human testing, but as complementary tools capable of supporting evidence-based design workflows, particularly for independent designers and smaller foundries with limited access to large-scale empirical testing.

Finally, the literature review argues that the integration of Variable Font technology, computational simulation, and empirical legibility research represents an important methodological shift within contemporary typography. By transforming typographic features into quantifiable and testable variables, researchers and designers are increasingly able to approach accessibility through systematic experimentation rather than subjective intuition alone. The review therefore establishes the theoretical and methodological foundation for investigating whether computational simulation models can function as reliable preliminary alternatives for evaluating typeface legibility in visually impaired contexts.

III.

Empirical

Work

6.

Methodology

6.1. Methodological Approach

As established in the introductory chapters, this study relied on a quantitative, quasi-experimental research design aimed at comparing human behavior with the VI-OCR pipeline simulations.

A quantitative, quasi-experimental design was adopted because the study aimed to compare measurable human performance outcomes with outputs generated by the VI-OCR simulation pipeline under visually impaired conditions. Since the aim of this study was to statistically compare two data sets, opting for a quantitative approach made the most sense. The study was considered quasi-experimental because participants could not be fully randomized across all human visual and behavioral variables, while the experimental conditions were instead implemented within a controlled but non-randomized setting.

The experimental procedure employed a modified staircase method adapted from the procedure described by Larson and Carter (2016). Rather than progressively varying exposure time, the present study implemented the staircase through typographic size reduction in order to evaluate legibility thresholds across four different typefaces under visually impaired conditions.

Participants were seated at a fixed viewing distance of 150 cm from an AOC 24B1H 23.6-inch monitor connected to an M4 MacBook Pro via HDMI. Stimuli were displayed at a resolution of 1920×1080 pixels with an effective pixel density of 94PPI. To ensure data integrity during the testing track, a static visual baseline image, serving as a fixation target, was introduced between the presentation of each consecutive trigram.

Prior to testing, monitor distance, resolution, and environmental conditions were verified to ensure consistency across sessions. The stimuli consisted of centrally displayed three-letter trigrams composed of unrelated characters (e.g., “pAe”) to minimize lexical prediction effects and isolate letter recognition performance. Participants completed two acclimation trials at 120pt before the beginning of the experimental sequence to ensure task comprehension. During each trial, a trigram was presented on-screen using one of the tested typefaces (Sitka, Luciole, Arial, or Times New Roman). Rather than reading the entire trigram, participants were asked to identify a randomly selected letter from the sequence, resulting in either a fail or a pass. This approach reduced contextual guessing and focused the task on isolated glyph recognition. The staircase sequence began at 120 pt and progressively decreased to 24 pt. Between stimuli presentations, as previously mentioned, a neutral transition screen (“eye cleanser”) was displayed to reduce visual persistence and fatigue effects.

The exact same images with the same trigrams, following the staircase procedure, were ran through 10 iterations of the VI-OCR pipeline with varying Hshift and Vshift user profiles. These images were exported at 1920×1080 pixels and 94PPI in order to ensure VI-OCR got the same input as human participants. The monitor distance in the VI-OCR pipeline was also maintained at 150cm.

The collected responses were subsequently compared against outputs generated by the VI-OCR simulation pipeline in order to evaluate how closely the computational model reproduced human legibility performance under visually impaired conditions.

As previously mentioned the experimental protocol evaluated four typefaces (Figure 24) selected to represent different approaches to legibility and accessibility in typographic design. For each tested typeface we decided to utilize the regular weight font file.

Sitka was included as a highly legibility-oriented serif typeface designed for screen reading and small-size rendering. Luciole was selected because it was specifically developed to support readers with visual impairments through increased character differentiation and open letterforms. As baseline references, the study also included Arial and Times New Roman, two widely used default system typefaces representative of conventional sans-serif and serif typographic standards respectively. In the case of these two typefaces we decided to utilize Monotype's/Microsoft's 5.01 version.

Sitka
Luciole
Arial
Times New Roman



Fig. 24. Comparison between each typeface.

The inclusion of these typefaces enabled comparisons between accessibility-oriented designs and commonly used digital reading fonts under simulated and real visually impaired reading conditions.

6.2. Research Site

Thanks to the generosity and the help of Tiago Silva, the empirical testing phase of this research was conducted at the ACAPO - *Associação dos Cegos e Amblíopes de Portugal* delegation in R. do Bonfim 215, 4300-069 Porto. The partnership with ACAPO, in the context of this study, was more than welcome as they are dedicated to building a more inclusive society by providing specialized rehabilitation, care, and guidance for visually impaired individuals.

The site was selected due to its accessibility infrastructure and its direct connection with individuals experiencing stable visual impairments, providing an appropriate environment for conducting legibility and perception studies involving visually impaired participants. Two testing sessions, on the 16th and the 23rd of April 2026, took place in a controlled indoor setting where viewing distance, monitor positioning, lighting conditions, and participant seating arrangements could be standardized across all trials. The research setup, as previously mentioned, consisted of an M4 MacBook Pro connected to an AOC 24B1H 23.6-inch monitor operating at a native resolution of 1920 × 1080 pixels and two reading materials, a magazine and a newspaper. The AOC 24B1H monitor was selected due to equipment availability and because its specifications are representative of commonly accessible consumer-grade displays, allowing the study conditions to more closely reflect typical everyday digital reading environments while still providing stable and standardized stimulus presentation.

Conducting the study within the ACAPO environment also facilitated participant recruitment and ensured that the experimental procedures could be carried out in a familiar and accessible context for the participants involved. Although the study was conducted across two rooms of different dimensions, the experimental setup remained standardized across both environments (Figure 25), maintaining consistent monitor positioning, participant seating distance, lighting conditions, and stimulus presentation arrangements throughout all testing sessions.

It is also worth mentioning that two colleagues of the Master's program, Jorge and Ana (R2), helped me (R1) in both of the testing sessions (Jorge on the 16th and Ana on the 23rd), their tasks consisted in keeping track of time, ensure participant distance to the monitor was maintained at 150cm and filling a google sheets file with notes and data whilst I conducted the protocol, guiding the participants and taking my own individual notes. Although they were essential as a part of the testing stage, I did not account for their personal laptops as part of the materials needed to run the protocol. These were used as per their own individual preference, as they could have also used the google sheets tables that needed to be filled, in a printed physical format that I had ready for them.

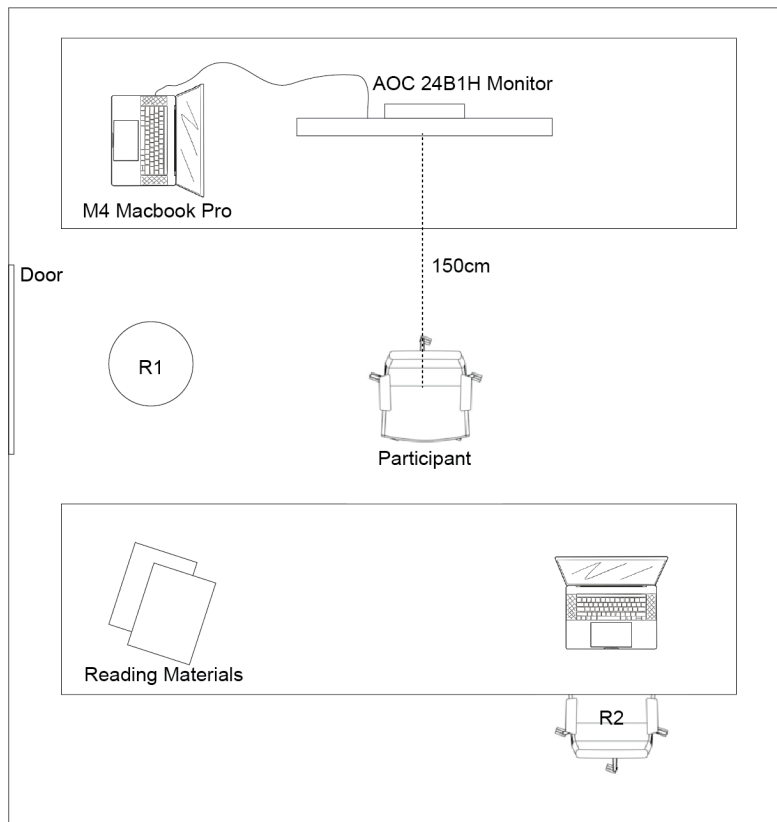


Fig.25. Example of Test Protocol Room Layout.

6.3. Participants and Sampling

As previously mentioned in other chapters the possible study group consists of individuals with visual impairments, excluding people who have vision worse than 6/12 and 6/60 in the better eye, that don't have a stable visual condition and that have associations with the research team.

After establishing contact with ACAPO through email and a phone call, Tiago Silva (their representative in the Porto delegation) kindly helped with the logistics behind the recruitment of the participants needed for the study and the testing in both sessions.

The human participants were recruited through ACAPO using a convenience sampling approach. We requested participants with stable visual conditions and levels of visual impairment within the thresholds required for our testing protocol. ACAPO then identified and provided eligible participants who were available to participate in the study.

The final testing was conducted with seven voluntary participants natural from the city of Porto, from Portuguese nationality, with ages between 24 and 71 years old, four biological female participants and three biological male participants. Out of the seven participants three were removed from the study, two cases because the participants had worse vision than our intended study threshold and one because ACAPO told us we unfortunately had run out of time and had to stop the testing protocol.

All participants had their information anonymized according to RGPD standard procedures, signing an informed consent form. For the purposes of the development of this study and this document, participants were assigned a

number following a standardized naming structure (P01, P02, P03, etc). The ten iterations of VI-OCR were also assigned a standardized naming structure (VI-OCR01, VI-OCR02, VI-OCR03, etc).

In the same way we removed participants because their vision was worse than our intended threshold we also had to remove three iterations of VI-OCR, with varying HShift and Vshift values since the “vision” of the simulation was also below the defined threshold. If we hadn’t removed these, it could negatively impact the analysis of the gathered data from our presented stimuli (Figure 26). The HShift and VShift values were selected from the list made available from Gao et al. (2025) research.

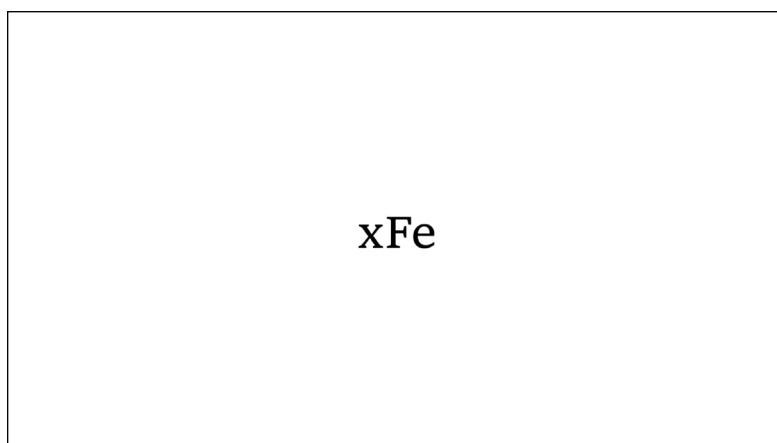


Fig.26. Example of presented Stimuli from the Test Protocol (see Appendix C).

P01, male, 71 years of age, felt ok and was well rested. His primary self-diagnosis consisted of high Myopia, retinal detachment, history of cataracts (removed), 65% of impairment/disability rate given by the Portuguese Government. He failed at the 10th line of the ETDRS test, the failed triplet in the Pelli-Robson test was ZSV - ran non-medically just to qualitatively give us more information to better understand his vision profile (this will be explained in the data collection instruments sub-chapter). In the first reading material (detailed in the data collection instruments sub-chapter) he had an average of 94 words read per minute. In the second material he had an average of 83 words read per minute, revealing more difficulty reading the second material. Used glasses with progressive lenses during testing.

P02, female, 43 years of age, felt ok and was well rested. Her primary self-diagnosis consisted of having multi-pathological eyes which included cataracts, nystagmus, strabismus, glaucoma, and other ocular pathologies. The participant reported better vision in the left eye, progressive deterioration in visual quality characterized by increasingly blurred and narrowed vision, stable prescription levels over time, good color perception, and the ability to recognize faces primarily with the left eye. She failed at the 5th line of the ETDRS test, the failed triplet in the Pelli-Robson test was CDN - ran non-medically just to qualitatively give us more information to better understand her vision profile (this will be explained in the data collection instruments sub-chapter). In the first reading material (detailed in the data collection instruments sub-chapter) she had an average of 68 words read per minute. In the second material she had an average of 42 words read per minute. Although it was still possible to get the participants average words per minute she struggled with reading the materials reporting she felt very tired. Used glasses with a prescription of 15.5 dioptres in the left eye and 14.5 dioptres in the right eye.

P03, female, 24 years of age, felt ok and was well rested. Her primary self-diagnosis consisted of being born premature, her visual condition was stable

but she reported everything close to her eyes is clear and everything far away is blurry, mentioned that at an average distance of over 2 meters she isn't able to identify anything. She reported having difficulty perceiving colors and that she uses more central vision as her peripheral vision is more compromised. The participant mentioned she has 10% usable vision. She failed at the 3rd line of the ETDRS test, the failed triplet in the Pelli-Robson test was VRS - ran non-medically just to qualitatively give us more information to better understand her vision profile (this will be explained in the data collection instruments sub-chapter). In the first reading material (detailed in the data collection instruments sub-chapter) she had an average of 18 words read per minute. In the second material she had an average of 24 words read per minute. The participant did not use any corrective equipment.

P04, male, 62 years of age, felt tired as he had gone to sleep late. His primary self-diagnosis consisted of reported vision exclusively in the right eye, with reduced contrast sensitivity. The participant stated being able to gradually adapt to backlighting conditions and preferred using dark mode on screens, as bright white backgrounds caused discomfort. The participant presented retinal problems, pathological myopia (approximately 18 diopters), and a history of retinal detachment in the left eye at the age of 36, resulting in complete vision loss in that eye. In 2001, retinal detachment was detected in the right eye and successfully treated surgically, allowing full visual recovery. In 2006, the participant was diagnosed with an epiretinal membrane, requiring photosensitive medication and laser treatment. The participant also reported blurred and fluctuating vision, a previous history of cataracts, and a current visual acuity of approximately 3 to 4 tenths. The participant used two pairs of glasses, one for near vision and another for distance vision, in addition to lenses with contrast-enhancing filters. The participant was eliminated as his usable vision was worse than the study threshold making him unable to complete the screening test and the staircase procedure of the trigrams.

P05, female, 29 years of age, felt ok and was well rested. The participant was eliminated as ACAPO told us we unfortunately had run out of time and had to stop the testing protocol.

P06, male, 61 years of age, felt ok and was well rested. The participant reported an ischemic stroke affecting the optic nerve, resulting in the ability to perceive visual stimuli without clearly distinguishing shapes. The participant stated that visual discrimination varied depending on fatigue levels, and that sunlight and fog negatively affected mobility and navigation. The participant was eliminated as his usable vision was worse than the study threshold making him unable to complete the screening test and the staircase procedure of the trigrams.

P07, female, 46 years of age, felt ok and was well rested. Her primary self-diagnosis consisted of only having usable peripheral vision only in the left eye. She failed at the 4th line of the ETDRS test, the failed triplet in the Pelli-Robson test was NHC - ran non-medically just to qualitatively give us more information to better understand her vision profile (this will be explained in the data collection instruments sub-chapter). She was not able to read any of the reading materials as they proved very difficult, she could identify colors and mentioned that the spacing is incredibly important for her. Used glasses during testing, although she mentioned they may be adjusted.

Regarding VI-OCR, the following user profiles were used in the ten iterations:

- VI-OCR01 - HShift = 4.1684 and VShift = 0.0575;
- VI-OCR02 - Hshift = 1.0 and VShift = 0.008;
- VI-OCR03 - HShift = 2.0 and VShift = 0.02;
- VI-OCR04 - HShift = 3.0 and VShift = 0.035;
- VI-OCR05 - HShift = 5.5 and VShift = 0.075;

VI-OCR06 - HShift = 7.0 and VShift = 0.09 (Removed because simulated vision was below the defined threshold);
 VI-OCR07 - HShift = 8.5 and VShift = 0.11 (Removed because simulated vision was below the defined threshold);
 VI-OCR08 - HShift = 10.0 and VShift = 0.125 (Removed because simulated vision was below the defined threshold);
 VI-OCR09 - HShift = 4.8 and VShift = 0.065;
 VI-OCR10 HShift = 6.2 and VShift = 0.085.

6.4. Data Collection Instruments

To achieve the primary objectives of the test protocol (Figure 27) and generate the needed data sets we used a main instrument (Google Sheets) divided into several instruments (tabs). The choice of these specific instruments over purely qualitative, quantitative or strictly clinical approaches allowed us to get a better understanding of the participants' difficulties and profiles. This allows for a direct, statistically rigorous comparison between both datasets.

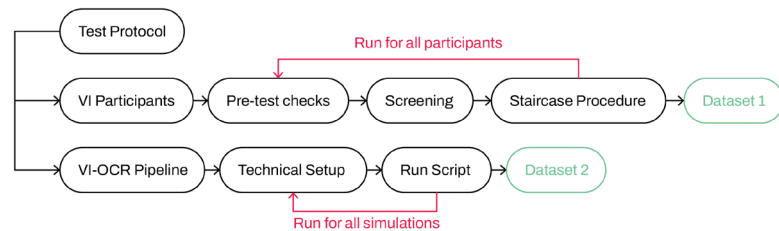


Fig.27. Summary of the Test Protocol.

Pre-testing checks

Description: Pre-check before starting the testing protocol with the goal of ensuring the environmental and participant conditions are normalized.

Objectives:

1. Ensure the participants physical distance to the AOC24B1H monitor is at 150cm;
2. Check if the participant is feeling well, if they are well rested and if they don't need to drink water or go to the bathroom before starting;
3. Explanation (reading out loud for the participant) and signing of the consent form;
4. Explanation of the testing procedure to the participant;
5. Gather information on biological gender and age of the participant.

Expected Results: Participants within 20 and 70 years old, well-rested and feeling good. The starting procedure runs well, participants understand the test protocol, and a fairly similar number of biological male and female participants. A few participants will be eliminated because their vision is unfortunately worse than our set threshold for this study.

Screening

Description: Screening of the participant's conditions through self-reported diagnosis and initial non-medical testing procedures utilizing two reading materials and EDTRS and Pelli-Robson charts. The reading materials were the most bought national newspaper, "Jornal de Notícias", from the 16th of April 2026 (Figure 28) and the Portuguese version of the National Geographic Magazine, Nº301, April 2026 (Figure 29) in physical format.

Iranianos alertam que cerco a portos é "prelúdio" para violar trégua e ameaçam suspender comércio pelo mar Vermelho

Gabriel Hansen
gabriel.hansen@jp.pt

IRÃO França e Reino Unido estão a preparar um plano para desbloquear o estreito de Ormuz que exclui, além do Irão, os Estados Unidos, revelou a imprensa norte-americana. Já Teerão considera que a operação militar norte-americana que impede o comércio através dos portos iranianos indica uma possível violação do cessar-fogo – e que está pronto para restringir o fluxo de navios no mar Vermelho também.

A proposta europeia será discutida amanhã pelo presidente francês, Emmanuel Macron, e o primeiro-ministro britânico, Keir Starmer, numa videoconferência, segundo o periódico "The Wall Street Journal", que revelou o plano. A missão será "puramente defensiva" e envol-

veria países "não beligerantes", excluindo-se então os EUA, Israel e o Irão, de acordo com o Palácio do Eliseu.

O jornal detalha que a operação consistirá na ajuda para a retirada dos navios presos no Golfo Pérsico, na remoção de minas marítimas colocadas pelo Irão e pela realização de escoltas militares regulares para proteção das embarcações comerciais. Tenta-se o apoio da Alemanha que, por sua vez, informa o diário, precisa de um mandato internacional – aprovado pelo Conselho de Segurança da ONU ou parte da missão da União Europeia "Aspides" alargada para além da questão dos hutis no mar Vermelho.

Mesmo que Starmer tenha dito que não "cederá" à pressão de Donald Trump, a possível rejeição do presidente norte-americano a tal ideia poderá ser o grande obstáculo para a sua execução. O líder da Casa Branca alegou ontem que a China estava "muito feliz" consigo por reabrir o estreito de Ormuz – que estava aberto para Pequim até ao bloqueio de Washington. "Estou a fazê-lo por eles também – e pelo

Mundo. Esta situação jamais se repetirá", escreveu.

O comandante militar iraniano Ali Abdollahi, por sua vez, declarou na televisão estatal que se constituirá um "prelúdio" para a violação do cessar-fogo caso os Estados Unidos continuem com o bloqueio e "criem insegurança para as embarcações comerciais e os petroleiros do Irão". "As poderosas Forças Armadas da República Islâmica não permitirão que quaisquer exportações ou importações continuem no Golfo Pérsico, no mar de Omã e no mar Vermelho", ameaçou.

SEM ESTENDER CESSAR-FOGO

As negociações prosseguem através do Paquistão, que continua a mediar as conversações. Apesar de um representante de Washington ter sublinhado à agência France-Presse que os EUA não concordaram formalmente com uma extensão do cessar-fogo, previsto para acabar na próxima semana, Trump mostra-se otimista. Em entrevista à estação Fox News, o conservador frisou que a guerra "está muito perto do fim".

urgindo a população a deslocar-se para norte do rio Zahrani – que já fica ao norte do rio Litani, considerado a divisão entre a parte meridional e o restante do país. Os israelitas atingiram mais de 200 alvos no Sul nas últimas 24 horas, alegando serem "infraestruturas terroristas do Hezbollah". Pelo menos dois veículos foram atingidos ainda nos arredores de Beirute. Já o grupo islamita anunciou que lançou dez rockets contra áreas no Norte de Israel – enquanto os hebraicos detetaram "aproximadamente 30 lançamentos". Também ontem, um deputado do Hezbollah disse que o grupo está a prevenir a tomada da cidade de Bint Jbeil, a cinco quilómetros da fronteira com Israel – e onde dez soldados israelitas ficaram feridos, na terça-feira.

Fig. 28. Reading Material from the "Jornal de Notícias" Newspaper.

Depois do declínio, na última década, registou-se um aumento do interesse pelos cafés de jazz por parte de jovens colecionadores de discos e melómanos. Numa época em que a música é um utilitário digital consumido através de auriculares, a visita aos *kissa* tornou-se popular entre os melómanos japoneses e uma espécie de peregrinação para ocidentais como eu. Todos procuramos uma oportunidade para experimentar a audição como um acto de veneração, apreciação e companheirismo.

A cultura *kissa* surgiu do *fandom* (sentimento de comunidade) partilhado. O primeiro café, Blackbird, abriu em 1929 perto da Universidade de Tóquio, inspirado pelo jazz norte-americano que chegara ao país recentemente através dos navios de cruzeiro ancorados em cidades portuárias como Yokohama. Durante a Segunda Guerra Mundial o governo japonês proibiu naturalmente a cultura norte-americana, mas o jazz ressurgiu na década de 1950 após o fim do conflito, com os discos trazidos dos Estados Unidos pelos militares norte-americanos destacados para as bases aéreas perto de Tóquio.

Os primeiros *kissa* tinham uma finalidade prática: até à década de 1960, poucos artistas de jazz fizeram digressões pelo Japão e os discos importados eram proibitivamente caros. Os proprietários dos *kissa* criaram coleções e tornaram-se educadores (*senseis*), apresentando o jazz a uma nova geração.

O renascimento dos *kissa* ganhou impulso depois de o álbum "Moanin'", de Art Blakey & Jazz Messengers, se tornar um sucesso inesperado (seguido pela célebre digressão de Blakey pelo Japão em Janeiro de 1961). Os cafés proliferaram em todo o país. O romancista Haruki Murakami escreveu os seus três primeiros romances enquanto geria o seu próprio bar, o Peter Cat, de 1974 a 1981. Sabendo-o, o "Auto-retrato do escritor enquanto corredor de fundo" ganha outra dimensão e profundidade.

Os *kissa* tradicionais apresentam jazz das décadas de 1950 e 1960; outros alternam entre o antigo e o moderno, misturando música brasileira, *french chanson* e até tango. O Café Lion foi fundado em 1926 para ouvir música clássica. Em vez de tocar faixas individuais, como

os DJ, a maioria dos proprietários de *kissa* toca discos inteiros. Alguns dos que conhecemos falaram sobre *kiki o yomu*, "ler o ar" ou discernir a atmosfera. "Olho primeiro para o rosto do cliente", disse Shoichi Suzuki, proprietário do Genius, em Tóquio, que colocou um disco do saxofonista Benny Golson durante a nossa visita.

Masahiro Goto, um crítico de jazz que fundou o elegante Eagle em 1967, não deixa nada ao acaso. Tem cadernos manuscritos com ordens precisas de músicas extraídas de sequências de quatro discos relacionados. "A banda desenhada japonesa tem quatro painéis... não é?", disse ele. "Eu entendo a estrutura de uma história". Masahiro publicou um livro de metodologia com o título "Jazz Music Selection Guide: Secrets to Listening to a 4-Album Set". "Não vendeu nada", disse ele com um sorriso.

Não há dois *kissa* iguais. Cada um é uma expressão da personalidade do seu proprietário. O Jazz in Rokudenashi, em Quioto, é um bar semelhante a uma caverna com uma estética *punk*, onde o baterista Naohisa Yokota põe a tocar Thelonious Monk enquanto os clientes conversam, fumam cigarros e bebem uísque. O Downbeat, fundado em Yokohama em 1956, agora gerido por Shuhei Yoshihisa de 40 anos, é um verdadeiro museu decorado com páginas amareladas da revista "Downbeat". Miyoko Hayashi (sem qualquer relação com Moeko, do café Masako) abriu o Jazz Spot Candy na cidade de Chiba, nos arredores de Tóquio, para homenagear os seus artistas favoritos, como Keith Jarrett e John Coltrane.

Os proprietários de *kissa* são por vezes designados como *otaku*, entusiastas com interesses obsessivos. Ao lado dos cafés de jazz, existem *kissa ongaku* (música) dedicados ao rock, à música clássica ou aos géneros brasileiros. Em Tóquio, um bar chamado Slow Hand toca apenas músicas de Eric Clapton de 1968 a 1972.

No fundo, todos partilham uma compreensão cultural adquirida com os veteranos dos *kissa*, mestres que estabeleceram os rituais que definem a experiência actual. O mais importante foi Hozumi Nakadaira, um fotógrafo que fundou o Dig em Tóquio em 1961. O Dig tornou-se

Fig. 29. Reading Material from the National Geographic Magazine.

The choice was due to the familiarity participants had with the type of materials used in their lifetime, the variety between both types of materials and the format itself as it allowed for a better understanding of the participants' difficulties without digital support. Participants were asked to read a text from both materials starting from the marking in bright pink, they were told they should read out loud as they normally would, stopping when needed. They were informed we would tell them when to stop after a while. They weren't told they just had a minute to read the material, this of course in order for them to both not rush and feel frustrated or tired.

Objectives:

1. Understand the visual impairments of each participant;
2. Run non-medical testing to better understand the vision profile of each participant;

Expected Results: Each participant will have different types of visual impairments, each one of them will report them as fairly stable. It will be fairly easy to understand the visual limitations of each participant, although they will vary a lot, even if participants have similar types of impairment. Participants will experience difficulty with both reading materials, with the *Jornal de Notícias* newspaper being easier to read than the National Geographic magazine. Participants will use some type of corrective equipment, such as glasses. Some participants will report difficulty with light in the room and the contrast of black text with white background on the monitor. Some participants will not be able to read the materials and in some extreme cases we will not be able to conduct the ETDRS and Pelli-Robson tests as the participants' usable vision will be outside of our threshold.

Staircase Procedure

Description: Primary part of the test protocol accessing the participants and VI-OCR's performance across 60 trigrams showed on the AOC24B1H monitor. The trigrams vary in size from 120pt to 24pt and typeface (Sitka, Luciole, Arial, Times New Roman). Participants are shown the trigram for a period of time (10 to 30s depending on severity of impairment) and asked in a random order if they can identify the left, middle or right letter of the trigram. This order seems random to the participant, but it is predetermined by us for all participants. VI-OCR 1 through 10 receives all the images, from 120pt to 24pt and outputs the images filtered and a text file identifying if the AI OCR recognized or not the content of the trigram, in this case the specific letter is irrelevant as VI-OCR outputs the entire trigram as a text string.

Objectives:

1. Gather data on the participants and VI-OCR's ability to pass or fail recognition for each of the trigrams;

Expected Results: It will be fairly easy to gather the required data from the participants although our sample size will not be very big since it is difficult to recruit enough visually impaired individuals to have a larger sample. The set-up process to make VI-OCR work will be problematic but we eventually will gather the data needed for the test. Comparing specific human participants against specific VI-OCR simulations will show in some cases a positive statistical correlation between both, proving that VI-OCR accurately simulated that specific individual's visual impairment. The Sans serif typefaces will be the best performers in both humans and VI-OCR tests. A few iterations of VI-OCR will have to be removed from the data analysis as the combination of Hshift and Vshift values will result in simulation where, like humans, VI-OCR is below our set threshold. Both the participants and VI-OCR will easily pass the

first four trigrams at 120pt but almost none of them is going to pass the trigrams set at 24pt, with recognition of the trigrams dropping drastically from 48pt onwards. In some typefaces both humans and VI-OCR will struggle to properly identify trigrams such as “dQl”, mistaking the “l” for the “i”, this will be the same for trigrams with glyphs that can be easily confused if not designed to have a very distinct construction. Most of the legibility research that has been done and presented in the literature review will be correct.

6.5. Ethical Considerations

This study was designed and conducted in strict accordance with established international ethical frameworks for research integrity, including the principles outlined by the Committee on Publication Ethics (COPE) regarding data integrity, authorship, and transparency, as well as institutional guidelines for academic archiving.

The research protocol involved the participation of visually impaired (VI) individuals. In alignment with standard ethical practices for human-centric research, all participants provided explicit, written informed consent prior to the commencement of the testing sessions. Participants received comprehensive verbal and written information regarding the study’s exact nature, the tasks required (the psychophysical staircase legibility test), and the overarching objectives of the research project. It was explicitly stated that participation was entirely voluntary and that individuals retained the right to withdraw from the study at any point during the process without any negative consequences or need for justification.

To safeguard the privacy and confidentiality of the participants, a strict protocol of data minimization and anonymization was enforced. All personal identifiers were stripped during the initial screening process. Human participants were assigned alphanumeric codes (e.g., P01 to P07) in all research records, analysis sheets, and final text, ensuring that individual clinical profiles (such as specific visual acuities or self-reported diagnoses) cannot be traced back to identity.

In compliance with open science practices and university regulations, the anonymized datasets generated during this study will be deposited in the Academic Research Bank (institutional repository). This ensures long-term preservation, academic transparency, and the verifiability of the research findings while maintaining absolute participant confidentiality.

A unique ethical dimension of this thesis involves the use of Generative AI and Computer Vision models to simulate human visual impairments (VI-OCR). This aspect of the methodology was handled with careful consideration of the following - Digital simulation can sometimes risk oversimplifying or misrepresenting the lived experiences of disabled communities. To mitigate this ethical concern, the qualitative post-test interviews specifically asked the VI participants for their perspectives on AI simulation tools. This approach ensures that the community’s voice actively audits and contextualizes the technical execution of the simulation. In accordance with COPE standards against data fabrication or falsification, all computational parameters (e.g., H_SHIFT and V_SHIFT filter settings) and AI outputs are recorded with absolute transparency in `TestProtocolData.xlsx` to prevent biased selective reporting and to ensure full reproducibility of the AI-human comparison.

6.6. Operationalization

As previously mentioned the test protocol followed a step by step logic in order to make testing as efficient and as result focused as possible. This process consisted of three core stages: Pre-Testing Checks, Screening and the Staircase Procedure.

After the participants entered the room they were welcomed and the pre-testing checks could begin, firstly the participants were informed about the research objectives and the details of the consent form (Figure 30).



CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)
Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 16 de Abri de 2026

Assinatura do Participante: _____

Assinatura do Investigador: _____

Fig.30. Signed Consent Form (see Appendix D).

After signing the consent form the participants were asked how they felt, if they were well rested and if they did not need a cup of water or to go to the bathroom, afterwards we asked their biological gender, their age and measured their distance to the AOC24B1H monitor, to ensure they were kept at a distance of 150cm.

Right after, we would begin screening the participants, this process would begin by quickly asking the participants what is their type of visual impairment, the type of difficulties they may feel in their day-to-day lives and if their visual condition was stable. Afterwards participants were asked to put on any type of corrective equipment they may use (e.g., glasses) for the remainder of the session.

Following this, participants were informed they would be reading, out-loud, texts from two reading materials, the newspaper and the magazine, starting from the bright pink marking on the page. Participants weren't told they would have only one minute to read the test materials, instead they were just told the researchers conducting the protocol would tell them when to stop.

The newspaper was first shown to all the participants followed by the magazine. The participants were then shown the ETDRS chart (Figure 31) at a distance of 150cm from the monitor, being asked to read each line until we accessed where they failed to read 3 letters. Afterwards participants were shown the Pelli-Robson Contrast Sensitivity Chart (Figure 32) at the same distance in the same monitor and they were asked to read each triplet top down, left to right until they failed two or more letters in a triplet. These procedures were done in a non-medical way as the goal was to get a qualitative rather than quantitative analysis of their vision to better understand their impairments. It is also worth noting that in all of the test protocol participants utilized both eyes.

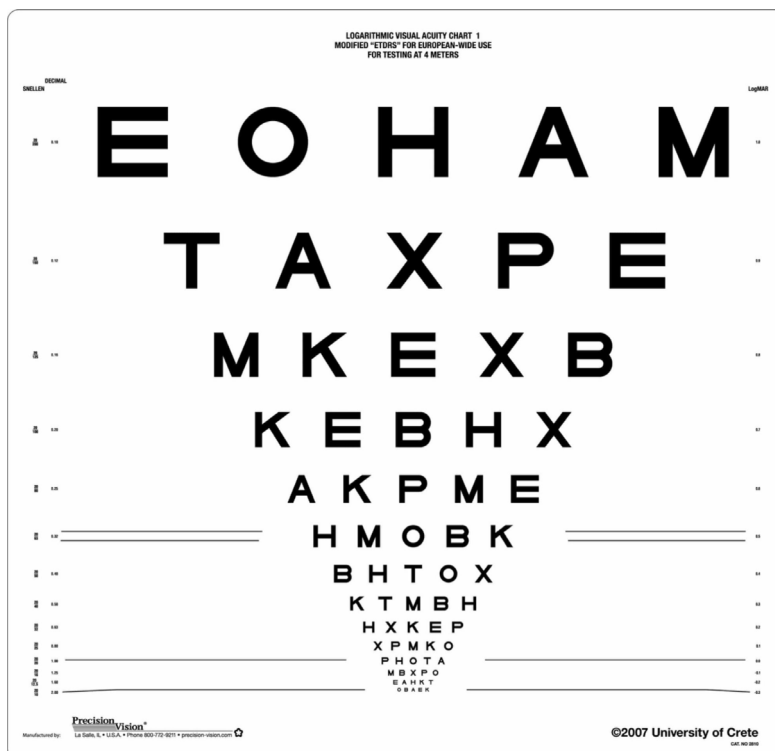


Fig.31. Precision Vision ETDRS Chart.



Fig.32. Precision Vision Pelli-Robson Contrast Sensitivity Chart.

The staircase procedure was conducted afterwards, the participants were informed trigrams would appear on the screen for a brief amount of time, and they would need to identify the left, right or middle letter according to what R1 (Figure 25) would ask. The order was random to them, but predetermined for all participants, so in the first trigram, for example, at 120pt, in the Sitka Regular Weight Font, all participants were asked if they could identify what was the letter in the middle. Between each trigram participants were shown a simple noise image (Figure 33) (different every time) to allow their visual system to return to baseline, mitigating afterimage effects. This process repeated itself from 120pt to 24pt across 60 trigrams. The participants were never told if they passed or failed a trigram.

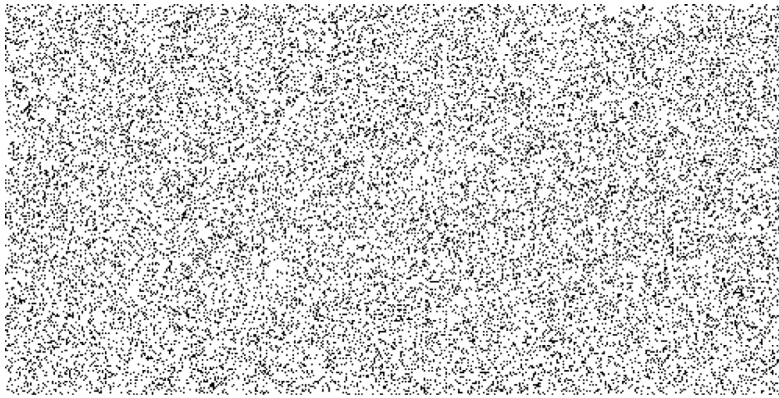


Fig.33. Example of used Noise Image to clear participants vision.

Afterwards, participants were asked four questions in a final interview:

- 1 - How are you feeling after the test? Do you feel tired?
- 2 - In your view, does the challenge of identifying these letters on screen accurately reflect the visual fatigue or barriers you encounter in your day to day life?
- 3 - What is your opinion on researchers using AI to simulate your vision to design better fonts?
- 4 - How much of an impact do you think this kind of research could have on your independence or quality of life?

For the simulations the testing is fairly similar, although run through VI-OCR instead of with human participants. The staircase procedure is exactly the same, with the same amount of trigrams being presented from 120pt to 24pt in the exact same typefaces. The resolution of the images that were input into VI-OCR was 1920x1080 at a resolution of 94PPI, respecting the same properties as the AOC24B1H monitor.

For the person utilizing the VI-OCR pipeline (our test version, with fixes to the original repository) the implementation is fairly easy to run, requiring only very basic knowledge of python.

The process starts by adding an images folder to the test_data folder, this images folder contains all the images that will be input into VI-OCR, without any filters, as they were presented to the human participants. These images must be exported according to the resolution and pixel density of the test monitor, in our case, 60 images, corresponding to 60 trigrams following a staircase, as previously mentioned, from 120pt to 24pt exported at 1920x1080 pixels at 94 PPI.

```

1  extract_and_save_filtered_fixed.py
2  Vi-OCR > extract_and_save_filtered_fixed.py > ...
3  5 import json
4  6 import time
5  7 from utils.filter import add_filter
6  8 import cv2
7  9 from tqdm import tqdm
8
9  10 # Setup API
10 11 api_key = 'your api key here'
11 12 genai.configure(api_key=api_key)
12 13
13 14 PROMPT = """What are all the visible sequences of 3 letters (trigrams) in this image? Please output each trigram in order from left to right and top to
14 15 bottom.
15 16 Important: Only output trigrams you are absolutely certain about. If text is blurry or unclear, DO NOT guess."""
16 17
17 18 # Vision parameters
18 19 H_SHIFT = 6.2
19 20 V_SHIFT = 0.085
20 21
21 22 # Create output directory for FILTERED IMAGES
22 23 filtered_images_dir = Path('outputs/filtered_images_Sub301')
23 24 filtered_images_dir.mkdir(parents=True, exist_ok=True)
24 25
25 26 print(f"{'='*60}")
26 27 print(f"Vi-OCR: Extracting trigrams AND saving filtered images")
27 28 print(f"{'='*60}")
28 29 print(f"Filtered images will be saved to: {filtered_images_dir}")
29 30 print(f"{'='*60}")
30 31
31 32 # Get ALL images
32 33 image_dir = Path('test_data/images')
33 34 all_images = sorted(image_dir.glob('*.jpg'))
34 35
35 36
36 37
37 38
38 39
39 40
40 41
41 42
42 43
43 44
44 45
45 46
46 47
47 48
48 49
49 50
50 51
51 52
52 53
53 54
54 55
55 56
56 57
57 58
58 59
59 60
60 61
61 62
62 63
63 64
64 65
65 66
66 67
67 68
68 69
69 70
70 71
71 72
72 73
73 74
74 75
75 76
76 77
77 78
78 79
79 80
80 81
81 82
82 83
83 84
84 85
85 86
86 87
87 88
88 89
89 90
90 91
91 92
92 93
93 94
94 95
95 96
96 97
97 98
98 99
99 100
100 101
101 102
102 103
103 104
104 105
105 106
106 107
107 108
108 109
109 110
110 111
111 112
112 113
113 114
114 115
115 116
116 117
117 118
118 119
119 120
120 121
121 122
122 123
123 124
124 125
125 126
126 127
127 128
128 129
129 130
130 131
131 132
132 133
133 134
134 135
135 136
136 137
137 138
138 139
139 140
140 141
141 142
142 143
143 144
144 145
145 146
146 147
147 148
148 149
149 150
150 151
151 152
152 153
153 154
154 155
155 156
156 157
157 158
158 159
159 160
160 161
161 162
162 163
163 164
164 165
165 166
166 167
167 168
168 169
169 170
170 171
171 172
172 173
173 174
174 175
175 176
176 177
177 178
178 179
179 180
180 181
181 182
182 183
183 184
184 185
185 186
186 187
187 188
188 189
189 190
190 191
191 192
192 193
193 194
194 195
195 196
196 197
197 198
198 199
199 200
200 201
201 202
202 203
203 204
204 205
205 206
206 207
207 208
208 209
209 210
210 211
211 212
212 213
213 214
214 215
215 216
216 217
217 218
218 219
219 220
220 221
221 222
222 223
223 224
224 225
225 226
226 227
227 228
228 229
229 230
230 231
231 232
232 233
233 234
234 235
235 236
236 237
237 238
238 239
239 240
240 241
241 242
242 243
243 244
244 245
245 246
246 247
247 248
248 249
249 250
250 251
251 252
252 253
253 254
254 255
255 256
256 257
257 258
258 259
259 260
260 261
261 262
262 263
263 264
264 265
265 266
266 267
267 268
268 269
269 270
270 271
271 272
272 273
273 274
274 275
275 276
276 277
277 278
278 279
279 280
280 281
281 282
282 283
283 284
284 285
285 286
286 287
287 288
288 289
289 290
290 291
291 292
292 293
293 294
294 295
295 296
296 297
297 298
298 299
299 300
300 301
301 302
302 303
303 304
304 305
305 306
306 307
307 308
308 309
309 310
310 311
311 312
312 313
313 314
314 315
315 316
316 317
317 318
318 319
319 320
320 321
321 322
322 323
323 324
324 325
325 326
326 327
327 328
328 329
329 330
330 331
331 332
332 333
333 334
334 335
335 336
336 337
337 338
338 339
339 340
340 341
341 342
342 343
343 344
344 345
345 346
346 347
347 348
348 349
349 350
350 351
351 352
352 353
353 354
354 355
355 356
356 357
357 358
358 359
359 360
360 361
361 362
362 363
363 364
364 365
365 366
366 367
367 368
368 369
369 370
370 371
371 372
372 373
373 374
374 375
375 376
376 377
377 378
378 379
379 380
380 381
381 382
382 383
383 384
384 385
385 386
386 387
387 388
388 389
389 390
390 391
391 392
392 393
393 394
394 395
395 396
396 397
397 398
398 399
399 400
400 401
401 402
402 403
403 404
404 405
405 406
406 407
407 408
408 409
409 410
410 411
411 412
412 413
413 414
414 415
415 416
416 417
417 418
418 419
419 420
420 421
421 422
422 423
423 424
424 425
425 426
426 427
427 428
428 429
429 430
430 431
431 432
432 433
433 434
434 435
435 436
436 437
437 438
438 439
439 440
440 441
441 442
442 443
443 444
444 445
445 446
446 447
447 448
448 449
449 450
450 451
451 452
452 453
453 454
454 455
455 456
456 457
457 458
458 459
459 460
460 461
461 462
462 463
463 464
464 465
465 466
466 467
467 468
468 469
469 470
470 471
471 472
472 473
473 474
474 475
475 476
476 477
477 478
478 479
479 480
480 481
481 482
482 483
483 484
484 485
485 486
486 487
487 488
488 489
489 490
490 491
491 492
492 493
493 494
494 495
495 496
496 497
497 498
498 499
499 500
500 501
501 502
502 503
503 504
504 505
505 506
506 507
507 508
508 509
509 510
510 511
511 512
512 513
513 514
514 515
515 516
516 517
517 518
518 519
519 520
520 521
521 522
522 523
523 524
524 525
525 526
526 527
527 528
528 529
529 530
530 531
531 532
532 533
533 534
534 535
535 536
536 537
537 538
538 539
539 540
540 541
541 542
542 543
543 544
544 545
545 546
546 547
547 548
548 549
549 550
550 551
551 552
552 553
553 554
554 555
555 556
556 557
557 558
558 559
559 560
560 561
561 562
562 563
563 564
564 565
565 566
566 567
567 568
568 569
569 570
570 571
571 572
572 573
573 574
574 575
575 576
576 577
577 578
578 579
579 580
580 581
581 582
582 583
583 584
584 585
585 586
586 587
587 588
588 589
589 590
590 591
591 592
592 593
593 594
594 595
595 596
596 597
597 598
598 599
599 600
600 601
601 602
602 603
603 604
604 605
605 606
606 607
607 608
608 609
609 610
610 611
611 612
612 613
613 614
614 615
615 616
616 617
617 618
618 619
619 620
620 621
621 622
622 623
623 624
624 625
625 626
626 627
627 628
628 629
629 630
630 631
631 632
632 633
633 634
634 635
635 636
636 637
637 638
638 639
639 640
640 641
641 642
642 643
643 644
644 645
645 646
646 647
647 648
648 649
649 650
650 651
651 652
652 653
653 654
654 655
655 656
656 657
657 658
658 659
659 660
660 661
661 662
662 663
663 664
664 665
665 666
666 667
667 668
668 669
669 670
670 671
671 672
672 673
673 674
674 675
675 676
676 677
677 678
678 679
679 680
680 681
681 682
682 683
683 684
684 685
685 686
686 687
687 688
688 689
689 690
690 691
691 692
692 693
693 694
694 695
695 696
696 697
697 698
698 699
699 700
700 701
701 702
702 703
703 704
704 705
705 706
706 707
707 708
708 709
709 710
710 711
711 712
712 713
713 714
714 715
715 716
716 717
717 718
718 719
719 720
720 721
721 722
722 723
723 724
724 725
725 726
726 727
727 728
728 729
729 730
730 731
731 732
732 733
733 734
734 735
735 736
736 737
737 738
738 739
739 740
740 741
741 742
742 743
743 744
744 745
745 746
746 747
747 748
748 749
749 750
750 751
751 752
752 753
753 754
754 755
755 756
756 757
757 758
758 759
759 760
760 761
761 762
762 763
763 764
764 765
765 766
766 767
767 768
768 769
769 770
770 771
771 772
772 773
773 774
774 775
775 776
776 777
777 778
778 779
779 780
780 781
781 782
782 783
783 784
784 785
785 786
786 787
787 788
788 789
789 790
790 791
791 792
792 793
793 794
794 795
795 796
796 797
797 798
798 799
799 800
800 801
801 802
802 803
803 804
804 805
805 806
806 807
807 808
808 809
809 810
810 811
811 812
812 813
813 814
814 815
815 816
816 817
817 818
818 819
819 820
820 821
821 822
822 823
823 824
824 825
825 826
826 827
827 828
828 829
829 830
830 831
831 832
832 833
833 834
834 835
835 836
836 837
837 838
838 839
839 840
840 841
841 842
842 843
843 844
844 845
845 846
846 847
847 848
848 849
849 850
850 851
851 852
852 853
853 854
854 855
855 856
856 857
857 858
858 859
859 860
860 861
861 862
862 863
863 864
864 865
865 866
866 867
867 868
868 869
869 870
870 871
871 872
872 873
873 874
874 875
875 876
876 877
877 878
878 879
879 880
880 881
881 882
882 883
883 884
884 885
885 886
886 887
887 888
888 889
889 890
890 891
891 892
892 893
893 894
894 895
895 896
896 897
897 898
898 899
899 900
900 901
901 902
902 903
903 904
904 905
905 906
906 907
907 908
908 909
909 910
910 911
911 912
912 913
913 914
914 915
915 916
916 917
917 918
918 919
919 920
920 921
921 922
922 923
923 924
924 925
925 926
926 927
927 928
928 929
929 930
930 931
931 932
932 933
933 934
934 935
935 936
936 937
937 938
938 939
939 940
940 941
941 942
942 943
943 944
944 945
945 946
946 947
947 948
948 949
949 950
950 951
951 952
952 953
953 954
954 955
955 956
956 957
957 958
958 959
959 960
960 961
961 962
962 963
963 964
964 965
965 966
966 967
967 968
968 969
969 970
970 971
971 972
972 973
973 974
974 975
975 976
976 977
977 978
978 979
979 980
980 981
981 982
982 983
983 984
984 985
985 986
986 987
987 988
988 989
989 990
990 991
991 992
992 993
993 994
994 995
995 996
996 997
997 998
998 999
999 1000
1000 1001
1001 1002
1002 1003
1003 1004
1004 1005
1005 1006
1006 1007
1007 1008
1008 1009
1009 1010
1010 1011
1011 1012
1012 1013
1013 1014
1014 1015
1015 1016
1016 1017
1017 1018
1018 1019
1019 1020
1020 1021
1021 1022
1022 1023
1023 1024
1024 1025
1025 1026
1026 1027
1027 1028
1028 1029
1029 1030
1030 1031
1031 1032
1032 1033
1033 1034
1034 1035
1035 1036
1036 1037
1037 1038
1038 1039
1039 1040
1040 1041
1041 1042
1042 1043
1043 1044
1044 1045
1045 1046
1046 1047
1047 1048
1048 1049
1049 1050
1050 1051
1051 1052
1052 1053
1053 1054
1054 1055
1055 1056
1056 1057
1057 1058
1058 1059
1059 1060
1060 1061
1061 1062
1062 1063
1063 1064
1064 1065
1065 1066
1066 1067
1067 1068
1068 1069
1069 1070
1070 1071
1071 1072
1072 1073
1073 1074
1074 1075
1075 1076
1076 1077
1077 1078
1078 1079
1079 1080
1080 1081
1081 1082
1082 1083
1083 1084
1084 1085
1085 1086
1086 1087
1087 1088
1088 1089
1089 1090
1090 1091
1091 1092
1092 1093
1093 1094
1094 1095
1095 1096
1096 1097
1097 1098
1098 1099
1099 1100
1100 1101
1101 1102
1102 1103
1103 1104
1104 1105
1105 1106
1106 1107
1107 1108
1108 1109
1109 1110
1110 1111
1111 1112
1112 1113
1113 1114
1114 1115
1115 1116
1116 1117
1117 1118
1118 1119
1119 1120
1120 1121
1121 1122
1122 1123
1123 1124
1124 1125
1125 1126
1126 1127
1127 1128
1128 1129
1129 1130
1130 1131
1131 1132
1132 1133
1133 1134
1134 1135
1135 1136
1136 1137
1137 1138
1138 1139
1139 1140
1140 1141
1141 1142
1142 1143
1143 1144
1144 1145
1145 1146
1146 1147
1147 1148
1148 1149
1149 1150
1150 1151
1151 1152
1152 1153
1153 1154
1154 1155
1155 1156
1156 1157
1157 1158
1158 1159
1159 1160
1160 1161
1161 1162
1162 1163
1163 1164
1164 1165
1165 1166
1166 1167
1167 1168
1168 1169
1169 1170
1170 1171
1171 1172
1172 1173
1173 1174
1174 1175
1175 1176
1176 1177
1177 1178
1178 1179
1179 1180
1180 1181
1181 1182
1182 1183
1183 1184
1184 1185
1185 1186
1186 1187
1187 1188
1188 1189
1189 1190
1190 1191
1191 1192
1192 1193
1193 1194
1194 1195
1195 1196
1196 1197
1197 1198
1198 1199
1199 1200
1200 1201
1201 1202
1202 1203
1203 1204
1204 1205
1205 1206
1206 1207
1207 1208
1208 1209
1209 1210
1210 1211
1211 1212
1212 1213
1213 1214
1214 1215
1215 1216
1216 1217
1217 1218
1218 1219
1219 1220
1220 1221
1221 1222
1222 1223
1223 1224
1224 1225
1225 1226
1226 1227
1227 1228
1228 1229
1229 1230
1230 1231
1231 1232
1232 1233
1233 1234
1234 1235
1235 1236
1236 1237
1237 1238
1238 1239
1239 1240
1240 1241
1241 1242
1242 1243
1243 1244
1244 1245
1245 1246
1246 1247
1247 1248
1248 1249
1249 1250
1250 1251
1251 1252
1252 1253
1253 1254
1254 1255
1255 1256
1256 1257
1257 1258
1258 1259
1259 1260
1260 1261
1261 1262
1262 1263
1263 1264
1264 1265
1265 1266
1266 1267
1267 1268
1268 1269
1269 1270
1270 1271
1271 1272
1272 1273
1273 1274
1274 1275
1275 1276
1276 1277
1277 1278
1278 1279
1279 1280
1280 1281
1281 1282
1282 1283
1283 1284
1284 1285
1285 1286
1286 1287
1287 1288
1288 1289
1289 1290
1290 1291
1291 1292
1292 1293
1293 1294
1294 1295
1295 1296
1296 1297
1297 1298
1298 1299
1299 1300
1300 1301
1301 1302
1302 1303
1303 1304
1304 1305
1305 1306
1306 1307
1307 1308
1308 1309
1309 1310
1310 1311
1311 1312
1312 1313
1313 1314
1314 1315
1315 1316
1316 1317
1317 1318
1318 1319
1319 1320
1320 1321
1321 1322
1322 1323
1323 1324
1324 1325
1325 1326
1326 1327
1327 1328
1328 1329
1329 1330
1330 1331
1331 1332
1332 1333
1333 1334
1334 1335
1335 1336
1336 1337
1337 1338
1338 1339
1339 1340
1340 1341
1341 1342
1342 1343
1343 1344
1344 1345
1345 1346
1346 1347
1347 1348
1348 1349
1349 1350
1350 1351
1351 1352
1352 1353
1353 1354
1354 1355
1355 1356
1356 1357
1357 1358
1358 1359
1359 1360
1360 1361
1361 1362
1362 1363
1363 1364
1364 1365
1365 1366
1366 1367
1367 1368
1368 1369
1369 1370
1370 1371
1371 1372
1372 1373
1373 1374
1374 1375
1375 1376
1376 1377
1377 1378
1378 1379
1379 1380
1380 1381
1381 1382
1382 1383
1383 1384
1384 1385
1385 1386
1386 1387
1387 1388
1388 1389
1389 1390
1390 1391
1391 1392
1392 1393
1393 1394
1394 1395
1395 1396
1396 1397
1397 1398
1398 1399
1399 1400
1400 1401
1401 1402
1402 1403
1403 1404
1404 1405
1405 1406
1406 1407
1407 1408
1408 1409
1409 1410
1410 1411
1411 1412
1412 1413
1413 1414
1414 1415
1415 1416
1416 1417
1417 1418
1418 1419
1419 1420
1420 1421
142
```

```

18 def add_filter(img,HShift,VShift,screen_reso = (1920,1080), screen_size = 24, camera = False, white_balance=False, resize=True):
19     v = screen_reso[1]
20     h = int(v*aspect_ratio)
21     charIm = cv2.resize(charIm,(h,v), interpolation = cv2.INTER_AREA)
22
23     # if image not taken by a camera therefore viewing angle depend on the viewing distance
24     if not camera:
25         # Calculate Viewing Angle
26         # PPI = sqrt((horizontal reso/width in inches)^2 + (vertical reso/height in inches)^2)
27         # PPI = PPI/2.54
28         PPI = np.sqrt((screen_reso[0]**2 + screen_reso[1]**2)/screen_size)
29         ppcm = PPI/2.54
30         PhysicalWidth = h/ppcm # physical width/height of the image on the screen (cm)
31         PhysicalHeight = v/ppcm # physical width/height of the image on the screen (cm)
32         distance=150 # Viewing distance in cm
33         vh = 2*math.atan(PhysicalWidth/(2*distance))*180/math.pi #horizontal visual angle of the image at the specified viewing distance
34         vw = 2*math.atan(PhysicalHeight/(2*distance))*180/math.pi #vertical visual angle of the image at the specified viewing distance
35         imgSize = vhwv #visual angle of the entire image at the specified viewing distance
36     else:
37         if v > h:
38             vw = 71
39             vh = 56
40         else:
41             vh = 71 # iPhone main camera horizontal field of view in degrees
42             vw = 56 # iPhone main camera vertical field of view in degrees
43         imgSize = vhwv # visual angle of the entire image
44
45     fx = np.arange(start=h/2, stop=h/2, step=1)
46     fy = fx/vh
47
48     # Processing images: 100% | 60/60 [03:50-00:00, 3.84s/it]
49
50     # COMPLETE!
51     # Filtered images saved to: outputs/filtered_images_Sub301
52     # Results saved to: outputs/complete_results_with_filtered_images.json
53     # Total filtered images saved: 60
54     # Total trigrams extracted: 60

```

Fig. 36. VI-OCR Pipeline Environment 3.

We also go to the filter.py script and change the distance to the physical distance between the participants and the monitor (Figure 36), in our case 150cm and also the line:

```
> def add_filter(img,HShift,VShift,screen_reso = (1920,1080), screen_size = 24, camera = False, white_balance=False, resize=True):
```

In this line we change the screen resolution and the screen size in inches, in our case 24in.

After opening our terminal we have to enter the virtual environment by running:

```
> source viocr_env/bin/activate
```

We can easily see we've entered the virtual environment successfully, as in our terminal, we will now have (viocr_env) before, in our case, hugo@Mac VI-OCR % and whatever we write after it.

After this we have to run:

```
> python3 extract_and_save_filtered_fixed.py
```

```

(viocr_env) hugo@Mac VI-OCR % python3 extract_and_save_filtered_fixed.py
/Users/hugo/Downloads/VI-OCR/VI-OCR/extract_and_save_filtered_fixed.py:2: FutureWarning:
All support for the 'google.generativeai' package has ended. It will no longer be receiving
updates or bug fixes. Please switch to the 'google.genai' package as soon as possible.
See README for more details.
https://github.com/google-gemini/deprecated-generative-ai-python/blob/main/README.md
import google.generativeai as genai
VI-OCR: Extracting trigrams AND saving filtered images
Filtered images will be saved to: outputs/filtered_images_Sub301

Processing images: 100% | 60/60 [03:50-00:00, 3.84s/it]

# COMPLETE!
# Filtered images saved to: outputs/filtered_images_Sub301
# Results saved to: outputs/complete_results_with_filtered_images.json
# Total filtered images saved: 60
# Total trigrams extracted: 60

```

Fig. 37. VI-OCR Pipeline Environment 4.

Whilst running the script the terminal will show a progress bar that eventually will reach 100% (Figure 37) giving us a "complete!" message and informing us where the filtered images and the results were outputted to.

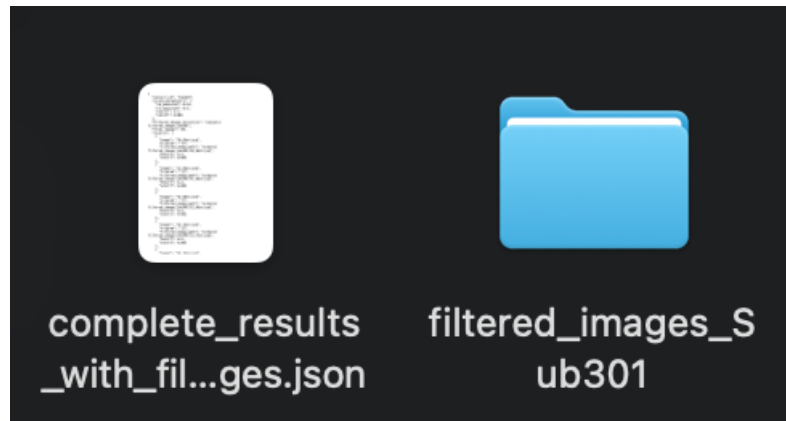


Fig.38. VI-OCR Outputs folder.

We should see a .json file and a folder with the original images with the filters applied to them in the VI-OCR/Outputs folder (Figure 38).

```

{
  "subject_id": "Sub301",
  "vision_parameters": {
    "va_measured": 0.54,
    "cs_measured": 0.9,
    "hshift": 4.1684,
    "vshift": 0.0575
  },
  "filtered_images_directory": "outputs/filtered_images_Sub301",
  "total_images": 60,
  "results": [
    {
      "image": "10_96pt.jpg",
      "trigram": "[jXh]",
      "filtered_image_path": "outputs/filtered_images_Sub301/10_96pt.jpg",
      "hshift": 4.1684,
      "vshift": 0.0575
    },
    {
      "image": "11_96pt.jpg",
      "trigram": "[dOl]",
      "filtered_image_path": "outputs/filtered_images_Sub301/11_96pt.jpg",
      "hshift": 4.1684,
      "vshift": 0.0575
    },
    {
      "image": "12_96pt.jpg",
      "trigram": "[wDk]",
      "filtered_image_path": "outputs/filtered_images_Sub301/12_96pt.jpg",
      "hshift": 4.1684,
      "vshift": 0.0575
    },
    {
      "image": "13_72pt.jpg",
      "trigram": "[gSp]",
      "filtered_image_path": "outputs/filtered_images_Sub301/13_72pt.jpg",
      "hshift": 4.1684,
      "vshift": 0.0575
    },
    {
      "image": "14_72pt.jpg",
      "trigram": "[tjh]",
      "filtered_image_path": "outputs/filtered_images_Sub301/14_72pt.jpg",
      "hshift": 4.1684,
      "vshift": 0.0575
    },
    {
      "image": "15_72pt.jpg",
      "trigram": "[ELA]",
      "filtered_image_path": "outputs/filtered_images_Sub301/15_72pt.jpg",
      "hshift": 4.1684,
      "vshift": 0.0575
    },
    {
      "image": "16_72pt.jpg",
      "trigram": "[]",
      "filtered_image_path": "outputs/filtered_images_Sub301/16_72pt.jpg",
      "hshift": 4.1684,
      "vshift": 0.0575
    },
    {

```

Fig.39. VI-OCR Example of .Json file OCR results for VI-OCR01.

The .json file (Figure 39) will contain a text string with information detailing if the VI-OCR simulation recognized or not the trigrams. If the simulation recognized the trigram it will output the trigram, for example, in the image above, in the 10th trigram at 96pt it recognize the letters [jXh], if the simulation doesn't recognize the trigram it will output a blank string [], such as

the 16th trigram at 72pt. Even though VI-OCR outputs trigrams that doesn't mean they are right, like human VI participants, VI-OCR can fail a trigram if it doesn't recognize a glyph correctly, such as by mistaking an "l" with "i", for example.

This exact process was ran 10 times for the 10 vision profiles (VI-OCR01 to VI-OCR10) across the 60 trigrams that were also shown to the participants, there was no attempt to run it multiple times in order to get better results for each vision profile. We ran each profile only once and analyzed the results from there in order to not taint test data. In our appendices you can see the filtered images for each of the VI-OCR simulations, compared to the original images.

After getting the test protocol fully implemented and obtaining all the data, the end result was a google sheets file, divided in tabs corresponding to the aforementioned sections (Pre-testing checks, Screening and Staircase procedure) to analyze and interpret.

The figure displays three screenshots of a Google Sheets spreadsheet, representing raw test protocol data. The top screenshot shows the 'Pre-test' tab, which includes columns for participant ID, name, and various test results. The middle screenshot shows the 'Screening' tab, which includes columns for participant ID, name, and various screening test results. The bottom screenshot shows the 'Staircase procedure' tab, which includes columns for participant ID, name, and various staircase procedure results. The data is organized into rows for each participant, with columns for different test conditions and results.

Fig.40. Raw Test Protocol Data (see Appendix E).

These of course aren't the easiest to read, the purpose of these being shown is to hopefully help contextualize the person reading this document, as in the next sub-chapter we will be breaking these down and showcasing how we got to our results. From the pre-test and the screening tabs we've already detailed our findings on previous sub-chapters, when describing the participants and their conditions.

It is easy to jump to conclusions in this stage of the process as certain patterns may start to emerge when looking at the staircase procedure tab. A few important things to note before going further is that, as described, P04, P05 and P06 are shown (orange columns), although, as previously mentioned they were excluded. VI-OCR06, VI-OCR07 and VI-OCR08 are also shown although they were excluded as well. All of these were excluded, as previously mentioned, because they did not fit the threshold we set for this study, if considered they would distort the data analysis.

For the more detail oriented you may also see that each typeface was tested the exact same amount of times, 15 trigrams per typeface to be exact and that there's an 11th VI-OCR simulation that hasn't been mentioned yet, this simulation named VI-OCR No Impairment is exactly what the name entails, a simulation with no filters applied (or without visual impairments) the purpose of this simulation was to test if VI-OCR was not "over applying" filters to the images by mistake which was a common problem we found when we were experimenting with our own self-built implementations of VI-OCR as it would apply some type of weird blur to images, even if the profile was that of a human person with normal vision. So, in summary, this simulation of VI-OCR was just used for testing purposes.

Last, but not least, it is also incredibly important to note that throughout this research to try and answer our secondary research questions, a few inquiries/questionnaires were made to people close to the research study (see Appendix F). At this point, a lot of the people that have been mentioned were from the type design industry, but there was also an attempt to reach academics and other equally important people. I'm deeply grateful for how this study allowed me to get out of my shell and not be afraid of reaching out as I do personally believe that it contributes to get an important, albeit qualitative understanding of how all involved parties feel about VI-OCR and how this type of technology could be implemented. It is also important to note, before we get to the breakdown of the results, that although I got in touch, got help and had a few important conversations with Qingying Gao and Dr.Yingzi Xiong about their research with VI-OCR, in no way or shape I'm biased towards it or its theoretical validity.

7.

Data Presentation and Findings

7.1. Results: Description and Analysis

The analysis of all of the data generated from the implemented test protocol was a fairly complex process.

Firstly, we started by creating a table, based on the Raw Staircase Procedure results, that replaced the Passed and Failure stats with 0 and 1 values. This was done in order to get numerical values for interpretation with mathematical formulas.

Typeface	Size in Points	Test	P01	P02	P03	P07	VI-OCR01	VI-OCR02	VI-OCR03	VI-OCR04	VI-OCR05	VI-OCR09	VI-OCR10
Sitka	120pt	xFe	1	1	0	1	1	1	1	1	1	0	1
Luciole	120pt	pAs	1	1	1	1	1	1	1	1	1	1	1
Arial	120pt	uKo	1	1	0	1	1	1	1	1	1	1	1
Times New Roman	120pt	vHr	1	1	0	0	1	1	1	1	1	1	1
Sitka	96pt	nZc	1	1	0	0	1	1	1	1	1	1	0
Sitka	96pt	fMu	1	0	0	0	1	1	1	1	1	0	0
Luciole	96pt	mEx	1	1	0	1	1	1	1	1	1	1	0
Luciole	96pt	rGy	1	1	0	1	1	1	1	1	1	1	1
Arial	96pt	bYl	1	0	0	0	1	1	1	1	1	1	1
Arial	96pt	jXh	1	1	1	0	1	1	1	1	0	0	0
Times New Roman	96pt	dQl	1	1	0	1	0	0	0	0	0	0	0
Times New Roman	96pt	wDk	1	1	0	1	1	1	1	1	0	1	1
Sitka	72pt	gSp	1	0	0	1	1	1	1	1	0	0	0
Sitka	72pt	lJh	1	0	0	0	1	1	1	0	0	0	0
Sitka	72pt	eLn	1	0	0	0	0	1	1	1	0	0	0
Luciole	72pt	oLn	1	1	0	0	0	1	1	1	0	0	0
Luciole	72pt	vQr	1	1	0	0	1	0	1	1	0	0	0
Luciole	72pt	aMt	1	1	0	0	1	1	1	1	0	0	0
Arial	72pt	kWz	1	1	0	0	0	1	1	1	0	0	1
Arial	72pt	aMt	1	1	0	0	1	1	1	1	0	0	0
Arial	72pt	oGy	1	0	0	0	1	1	1	1	0	0	0
Times New Roman	72pt	cPl	1	0	0	0	1	1	1	1	0	0	0
Times New Roman	72pt	sRd	1	1	0	0	0	1	1	1	0	0	0
Times New Roman	72pt	rVq	1	0	0	0	1	1	1	1	0	0	0
Sitka	48pt	vQl	1	0	0	0	0	0	0	0	0	0	0
Sitka	48pt	zOm	1	0	0	0	0	1	1	0	0	0	0
Sitka	48pt	iNb	1	0	0	0	0	0	1	0	0	0	0
Luciole	48pt	uKo	1	0	0	0	0	1	0	0	0	0	0
Luciole	48pt	yTf	1	0	0	0	0	1	1	1	0	0	0
Luciole	48pt	dQl	1	0	0	0	0	0	0	0	0	0	0
Arial	48pt	pUg	1	0	0	0	0	1	1	0	0	0	0
Arial	48pt	nZc	1	0	0	0	0	0	1	0	0	0	0
Arial	48pt	gSp	1	0	0	0	0	1	1	0	0	0	0
Times New Roman	48pt	hBe	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	48pt	xFe	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	48pt	kWz	1	0	0	0	0	0	1	1	0	0	0
Sitka	36pt	rVq	1	0	0	0	0	0	0	0	0	0	0
Sitka	36pt	wDk	1	0	0	0	0	1	0	0	0	0	0
Sitka	36pt	jXh	1	0	0	0	0	0	0	0	0	0	0
Luciole	36pt	sRd	1	0	0	0	0	1	0	0	0	0	0
Luciole	36pt	bYl	1	0	0	0	0	1	0	0	0	0	0
Luciole	36pt	fMu	1	0	0	0	0	0	1	0	0	0	0
Arial	36pt	eLn	1	0	0	0	0	1	0	0	0	0	0
Arial	36pt	mEx	1	0	0	0	0	1	1	0	0	0	0
Arial	36pt	cPl	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	36pt	aMt	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	36pt	pAs	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	36pt	uKo	1	0	0	0	0	0	0	0	0	0	0
Sitka	24pt	dQl	0	0	0	0	0	0	0	0	0	0	0
Sitka	24pt	oGy	0	0	0	0	0	0	0	0	0	0	0
Sitka	24pt	pUg	1	0	0	0	0	0	0	0	0	0	0
Luciole	24pt	kSo	1	0	0	0	0	0	0	0	0	0	0
Luciole	24pt	zOm	1	0	0	0	0	0	0	0	0	0	0
Luciole	24pt	nZc	1	0	0	0	0	0	0	0	0	0	0
Arial	24pt	xFe	1	0	0	0	0	0	0	0	0	0	0
Arial	24pt	hBe	0	0	0	0	0	0	0	0	0	0	0
Arial	24pt	vHr	0	0	0	0	0	0	0	0	0	0	0
Times New Roman	24pt	lJh	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	24pt	vQl	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	24pt	sRd	0	0	0	0	0	0	0	0	0	0	0

Fig.41. Edited Table showcasing the results from the Staircase Procedure.

The result was the table presented in Figure 41, showcasing either Passing or Failure (1 or 0) for each of the 60 trigrams across human participants and the VI-OCR simulations (excluding the ones we've eliminated previously).

Typeface - Sitka	Size in Points	Test	P01	P02	P03	P07	VI-OCR01	VI-OCR02	VI-OCR03	VI-OCR04	VI-OCR05	VI-OCR09	VI-OCR10
Sitka	120pt	xFe	1	1	0	1	1	1	1	1	1	0	1
Sitka	96pt	nZc	1	1	0	0	1	1	1	1	1	1	0
Sitka	96pt	fMu	1	0	0	0	1	1	1	1	0	0	0
Sitka	72pt	gSp	1	0	0	1	1	1	1	1	0	0	0
Sitka	72pt	lJh	1	0	0	0	1	1	1	0	0	0	0
Sitka	72pt	eLn	1	0	0	0	0	1	1	1	0	0	0
Sitka	48pt	vQl	1	0	0	0	0	0	0	0	0	0	0
Sitka	48pt	zOm	1	0	0	0	0	1	1	0	0	0	0
Sitka	48pt	iNb	1	0	0	0	0	0	1	0	0	0	0
Sitka	36pt	rVq	1	0	0	0	0	0	0	0	0	0	0
Sitka	36pt	wDk	1	0	0	0	0	1	0	0	0	0	0
Sitka	36pt	jXh	1	0	0	0	0	0	0	0	0	0	0
Sitka	24pt	dQl	0	0	0	0	0	0	0	0	0	0	0
Sitka	24pt	oGy	0	0	0	0	0	0	0	0	0	0	0
Sitka	24pt	pUg	1	0	0	0	0	0	0	0	0	0	0
Typeface - Luciole	Size in Points	Test	P01	P02	P03	P07	VI-OCR01	VI-OCR02	VI-OCR03	VI-OCR04	VI-OCR05	VI-OCR09	VI-OCR10
Luciole	120pt	pAs	1	1	1	1	1	1	1	1	1	1	1
Luciole	96pt	mEx	1	1	0	1	1	1	1	1	1	1	0
Luciole	96pt	rGy	1	1	0	1	1	1	1	1	1	1	1
Luciole	72pt	oLn	1	1	0	0	0	1	1	1	0	0	0
Luciole	72pt	vQr	1	1	0	0	1	0	1	1	0	0	0
Luciole	72pt	aMt	1	1	0	0	1	1	1	1	0	0	0
Luciole	48pt	uKo	1	0	0	0	0	1	0	0	0	0	0
Luciole	48pt	yTf	1	0	0	0	0	1	1	1	0	0	0
Luciole	48pt	dQl	1	0	0	0	0	0	0	0	0	0	0
Luciole	36pt	sRd	1	0	0	0	0	1	0	0	0	0	0
Luciole	36pt	bYi	1	0	0	0	0	1	0	0	0	0	0
Luciole	36pt	fMu	1	0	0	0	0	0	1	0	0	0	0
Luciole	24pt	kSo	1	0	0	0	0	0	0	0	0	0	0
Luciole	24pt	zOm	1	0	0	0	0	0	0	0	0	0	0
Luciole	24pt	nZc	1	0	0	0	0	0	0	0	0	0	0
Typeface - Arial	Size in Points	Test	P01	P02	P03	P07	VI-OCR01	VI-OCR02	VI-OCR03	VI-OCR04	VI-OCR05	VI-OCR09	VI-OCR10
Arial	120pt	uKo	1	1	0	1	1	1	1	1	1	1	1
Arial	96pt	bYi	1	0	0	0	1	1	1	1	1	1	1
Arial	96pt	jXh	1	1	1	0	1	1	1	1	0	0	0
Arial	72pt	kWz	1	1	0	0	0	1	1	1	0	0	1
Arial	72pt	aMt	1	1	0	0	1	1	1	1	0	0	0
Arial	72pt	oGy	1	0	0	0	1	1	1	1	0	0	0
Arial	48pt	pUg	1	0	0	0	0	1	1	1	0	0	0
Arial	48pt	nZc	1	0	0	0	0	0	1	0	0	0	0
Arial	48pt	gSp	1	0	0	0	0	1	1	0	0	0	0
Arial	36pt	eLn	1	0	0	0	0	1	0	0	0	0	0
Arial	36pt	mEx	1	0	0	0	0	1	1	0	0	0	0
Arial	36pt	cPl	1	0	0	0	0	0	0	0	0	0	0
Arial	24pt	xFe	1	0	0	0	0	0	0	0	0	0	0
Arial	24pt	hBe	0	0	0	0	0	0	0	0	0	0	0
Arial	24pt	vHr	0	0	0	0	0	0	0	0	0	0	0
Typeface - TNR	Size in Points	Test	P01	P02	P03	P07	VI-OCR01	VI-OCR02	VI-OCR03	VI-OCR04	VI-OCR05	VI-OCR09	VI-OCR10
Times New Roman	120pt	vHr	1	1	0	0	1	1	1	1	1	1	1
Times New Roman	96pt	dQl	1	1	0	1	0	0	0	0	0	0	0
Times New Roman	96pt	wDk	1	1	0	1	1	1	1	1	0	1	1
Times New Roman	72pt	cPl	1	0	0	0	1	1	1	1	0	0	0
Times New Roman	72pt	sRd	1	1	0	0	0	1	1	1	0	0	0
Times New Roman	72pt	rVq	1	0	0	0	1	1	1	1	0	0	0
Times New Roman	48pt	hBe	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	48pt	xFe	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	48pt	kWz	1	0	0	0	0	0	1	1	0	0	0
Times New Roman	36pt	aMt	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	36pt	pAs	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	36pt	uKo	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	24pt	lJh	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	24pt	vQl	1	0	0	0	0	0	0	0	0	0	0
Times New Roman	24pt	sRd	0	0	0	0	0	0	0	0	0	0	0

Fig.42. Edited Tables showcasing the Results per each typeface.

Besides this we've also made 4 other tables (Figure 42) that show the same results from Figure 41, but now divided for each of the tested typefaces. So in the Main table we have a total of 60 results per column and in the tables for each typeface it shows 15 results per column, essentially confirming what we've mentioned previously.

If we take another look at the main results table (Figure 41). The first trigram 'xFe' was shown at 120pt, in the Sitka typeface. P01, P02, P07, VI-OCR01, VI-OCR02, VI-OCR03, VI-OCR04, VI-OCR05 and VI-OCR10 all recognized the trigram with only P03 and VI-OCR09 failing to recognize this specific trigram, in this specific typeface, in this specific size in points.

Analysis P2							
P2	P2-VIOCR01	P2-VIOCR02	P2-VIOCR03	P2-VIOCR04	P2-VIOCR05	P2-VIOCR09	P2-VIOCR10
Both Pass	12	14	15	15	7	7	7
Both Fail	37	26	27	35	43	43	43
Human Pass/Sim Fail	4	2	1	1	9	9	9
Sim Pass/Human Fail	7	18	17	9	1	1	1
Total	60	60	60	60	60	60	60
Observed Agreement	0.817	0.667	0.700	0.833	0.833	0.833	0.833
Failing By Chance	0.084	0.142	0.142	0.107	0.036	0.036	0.036
Passing By Chance	0.501	0.342	0.342	0.440	0.636	0.636	0.636
Expected Agreement	0.586	0.484	0.484	0.547	0.671	0.671	0.671
Cohen Kappa	0.558	0.353	0.418	0.632	0.493	0.493	0.493
Discordant Total	11	20	18	10	10	10	10
x2	0.818	12.800	14.222	6.400	6.400	6.400	6.400
P Value	0.366	0.000	0.000	0.011	0.011	0.011	0.011

Fig.43. Example of comparative mathematical analysis between P02 and each of the VI-OCR simulations.

After this process we compared the results of each of the human participants against the results of each of the VI-OCR simulations (Figure 43) to try and understand which VI-OCR simulation performed better at simulating the vision of that specific visually impaired individual. Mathematically this was done by calculating the Cohen's Kappa and the P-value for each case.

Cohen's Kappa is a robust statistic designed to measure inter-rater agreement for categorical data (such as our binary 1 for Pass and 0 for Failure). While simple percentage agreement just calculates how often two raters get the same result, Cohen's Kappa goes a step further by calculating how much they agree above what would be expected by pure random guessing.

The statistical formula is defined as:

$$\kappa = (p_o - p_e) / (1 - p_e).$$

As noted in the formula for calculating Cohen's Kappa, p_o is the observed agreement, and p_e is the expected agreement. The observed agreement is the raw, actual percentage of trials where both raters gave the exact same answer. The expected agreement represents the probability that the two raters would agree with each other purely by random chance or accident.

p_o is calculated by adding the values for "Both Pass" and "Both Fail", these coming from the zeros and ones of the Main table in Figure 41. and then dividing that value by the total of tests. In the case of P2-VIOCR01, $p_o = (12+37)/60 = 0.817$.

p_e is calculated by adding the values of the Failing By Chance and the Passing by Chance results. In this case $p_e = 0.084 + 0.501 = 0.586$. Failing by Chance is calculated by adding the results of Both Pass and Human Pass/Sim Fail divided by the total, multiplied by the sum of Both Pass and Sim Pass/Human Fail divided by the total. So, in our case for P2-VIOCR01, Failing by Chance = $((12+4)/60 * (12+7)/60) = 0.084$.

Passing by chance follows a similar formula, this time adding the values for Both Fail and Sim Pass/Human Fail, dividing them by the total and multiplying them by the sum of Both Fail and Human Pass/Sim Fail, divided by the total. So, in our case for P2-VIOCR01, Passing By Chance = $((7+37)/60 * (4+37)/60) = 0.501$.

Calculating Cohen's Kappa from here is just a matter of applying the previously mentioned formula, so, for P2-VIOCR01, $k = (0.817 - 0.586) / (1 - 0.586) = 0.558$.

The P-value (Probability value) is a directional metric used in inferential statistics to determine if an observed result happened by random chance, or if it represents a meaningful pattern. In our data, the P-value was calculated using a Chi-Square (χ^2) test of independence mapped across the discordant totals (the

instances where the human and machine disagreed: Human Pass/Sim Fail and Sim Pass/Human Fail).

χ^2 was calculated according to the following formula:

$$\chi^2 = (b-c)^2 / (b+c)$$

Using our naming conventions, the formula for Chi-Squared would be $\chi^2 = (\text{HumanPassSimFail} - \text{SimPassHumanFail})^2 / (\text{HumanPassSimFail} + \text{SimPassHumanFail})$. So, in our case, for P2-VIOCR01, $\chi^2 = (4-7)^2 / (4+7) = 0.818$. For calculating the P-Value, we take our calculated Chi-Square value and determine its significance using a right-tailed Chi-Square distribution with 1 degree of freedom (1), in google sheets this can be easily done by doing `=CHISQ.DIST.RT( $\chi^2$ ,1)`. In our case, for P2-VIOCR01, we did `=CHISQ.DIST.RT(0.818,1)`, resulting in a P-Value of 0.366.

For analyzing Cohen's Kappa there are several different levels of alignment, 0 (Poor): The VI-OCR simulation and the human participant have absolutely zero meaningful alignment. Their agreement can be explained entirely by pure random luck. 0.01 – 0.20 (Slight): There is a tiny trace of structural overlap, but the simulation fails to capture the true reading limits of the participant. 0.21 – 0.40 (Fair): The model and the human start showing general trends of matching, but it is still too weak to rely on as a predictive tool. 0.41 – 0.60 (Moderate): A solid, acceptable level of structural overlap. The machine vision is successfully matching a significant portion of the human's specific pass/fail breaking points across the trigrams. 0.61 – 0.80 (Substantial): An excellent, highly accurate match. The simulation is operating as a near-perfect proxy, mirroring the human's eye limitations with incredible precision. 0.81 – 1.00 (Almost Perfect): The human and the simulation behave almost identically on virtually every single one of the 60 tested trigrams.

For analyzing the P-Value, a p-value < 0.05 is “bad” for simulation accuracy, meaning that the simulation differs significantly from the participant and that the VI-OCR simulation may not accurately reproduce that participant's visual behavior. On the other hand, a P-value > 0.05, is “good” because we fail to detect a significant difference, meaning the simulation behaves similarly to the participant, which supports the idea that that specific VI-OCR simulation presented no discernable differences in expected behaviour in recognition from the human participants.

While 0.05 is the standard scientific threshold, it is an arbitrary convention rather than an absolute rule of truth. For example, a p-value of 0.051 and 0.049 provide nearly identical evidence, but one is traditionally labeled “significant” and the other “not significant”.

For the previously mentioned example, P2-VIOCR01, demonstrated moderate agreement ($\kappa = 0.558$) between the participant and the VI-OCR simulation, while the non-significant p-value ($p = 0.366$) suggests no statistically detectable difference in overall performance. Together, these findings indicate that the simulation approximates participant behavior reasonably well, though not perfectly.

Although our sample is small to dive in exactly which typeface behaved best or what exact parameters affected trigram legibility at specific sizes, we developed a few extra tables to try and reach a few conclusions (Figure 44), minding that these aren't 100% set in stone and definitely would need to be corroborated with further research.

Participants	Sitka	Luciole	Arial	Times New Rom	Better Performer
P01	13	15	13	14	Luciole
P02	2	6	4	4	Luciole
P03	0	1	1	0	Luciole/Arial
P07	2	3	1	2	Luciole
VIOCR-01	5	5	5	4	Luciole/Arial/Sitka
VIOCR-02	8	9	10	5	Arial
VIOCR-03	8	8	10	6	Arial
VIOCR-04	5	7	6	6	Luciole
VIOCR-05	2	3	2	1	Luciole
VIOCR-09	1	3	2	2	Luciole
VIOCR-10	1	2	3	2	Arial
Delta	Sitka	Luciole	Arial	Times New Roman	
P01-VIOCR-01	8	10	8	10	
P01-VIOCR-02	5	6	3	9	
P01-VIOCR-03	5	7	3	8	
P01-VIOCR-04	8	8	7	8	
P01-VIOCR-05	11	12	11	13	
P01-VIOCR-09	12	12	11	12	
P01-VIOCR-10	12	13	10	12	
P02-VIOCR-01	-3	1	-1	0	
P02-VIOCR-02	-6	-3	-6	-1	
P02-VIOCR-03	-6	-2	-6	-2	
P02-VIOCR-04	-3	-1	-2	-2	
P02-VIOCR-05	0	3	2	3	
P02-VIOCR-09	1	3	2	2	
P02-VIOCR-10	1	4	1	2	
P03-VIOCR-01	-5	-4	-4	-4	
P03-VIOCR-02	-8	-8	-9	-5	
P03-VIOCR-03	-8	-7	-9	-6	
P03-VIOCR-04	-5	-6	-5	-6	
P03-VIOCR-05	-2	-2	-1	-1	
P03-VIOCR-09	-1	-2	-1	-2	
P03-VIOCR-10	-1	-1	-2	-2	
P07-VIOCR-01	-3	-2	-4	-2	
P07-VIOCR-02	-6	-6	-9	-3	
P07-VIOCR-03	-6	-5	-9	-4	
P07-VIOCR-04	-3	-4	-5	-4	
P07-VIOCR-05	0	0	-1	1	
P07-VIOCR-09	1	0	-1	0	
P07-VIOCR-10	1	1	-2	0	
MAE (Mean Absolute Error)	4.678571429	4.75	4.821428571	4.428571429	
MSB (Mean Signed Deviation)	-0.03571428571	0.9642857143	-0.6785714286	1.285714286	

Fig.44. Comparison of the performance for each of the test typefaces.

The first table scores how many trigrams were recognized for each typeface for each of the individual participants. So, in the case of Sitka, P01 recognized a total of 13 trigrams, and P02 only recognized 2 trigrams.

The second table, built from the first table, shows the delta between the human participants and each VI-OCR simulation. So, comparing P01-VIOCR-01, we get our delta = 13-5=8. At the end of the table we calculated the MAE (Mean Absolute Error) and the MSD (Mean Signed Deviation). Lower MAE is better because it means that across all testing blocks, the pipeline's final scores were mathematically closer and more stable to the real human participant scores. The individual trial gaps were small, making the simulation precise. Higher MAE is worse because it means that the pipeline's performance was highly erratic compared to the humans. The simulation scores fluctuated wildly from being way too high in some conditions to way too low in others. An MSD near 0 means the simulation pipeline has zero overall systemic bias. Even if there were individual errors (tracked by MAE), the instances where VI-OCR over-performed perfectly balanced out the instances where it under-performed. The further below zero the score goes, the more the VI-OCR pipeline over-performs compared to humans. The simulation filters are letting too much readable information get through to the OCR engine compared to what a real human eye can resolve under those same simulated conditions. The

further above zero the score goes, the more the VI-OCR pipeline under-performs compared to humans. The simulation parameters are degrading the text too aggressively, making it unreadable to the simulations OCR, even though a real human with low vision correctly identified the trigram.

With the calculation procedures and evaluation metrics now established, the results can be examined. To maintain consistency with the study's objectives, the findings are presented according to each research question and supported by the relevant tables and figures.

7.1.1. Primary Research Question

The primary objective of this study was to determine to what extent the quantitative legibility metrics generated by the VI-OCR computational simulations reliably correlate with data obtained from human participants. To empirically validate the core hypothesis, that the VI-OCR pipeline can serve as a scientifically backed way to simulate impaired human vision in legibility testing, the outcomes were cross-examined, following the previously mentioned methodology.

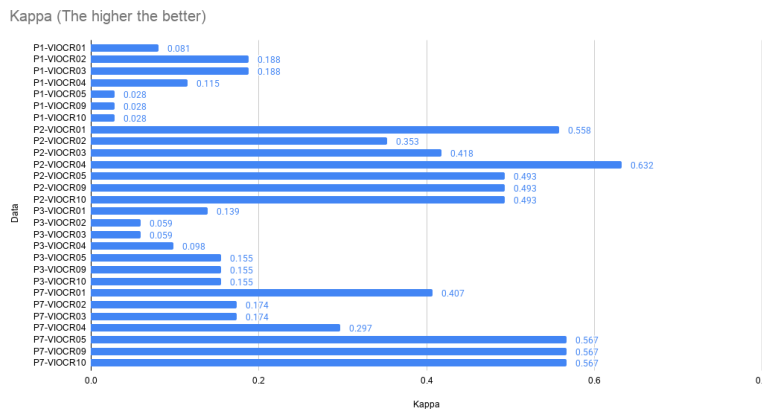


Fig.45. Comparison of Kappa results for each participant and VI-OCR simulation combination.

Using Cohen's Kappa, following the previously mentioned procedure and comparing the results across all participants and simulations, we reveal a few distinct performance situations that characterize the VI-OCR simulation pipeline (Figure 45).

Out of all of the crossed data between the VI-OCR simulations and the human participants, P2-VIOCR01, P2-VIOCR02, P2-VIOCR03, P2-VIOCR05, P2-VIOCR09, P2-VIOCR10, P7-VIOCR01, P7-VIOCR05, and P7-VIOCR09 show moderate agreement, this means that even after mathematically stripping away all occurrences of lucky coincidences, there is a distinct, measurable alignment between how the human eye and the VI-OCR pipeline processed the input images. Although viable, these combinations show that the simulations aren't still identical doubles, rather, it proves that the simulation pipeline is firmly on the right track. It demonstrates a mathematically sound correlation, showing that the applied filters and the character recognition engine are authentically mirroring a human subject's visual impairment profile rather than behaving erratically.

It is important to note that our analysis reveals that P2-VIOCR04 has the highest value of $k=0.632$, showing substantial agreement, which means that the specific filtering applied in the VI-OCR04 simulation operates as a high-fidelity digital twin of this individual participant's visual constraints

(P02), mirroring their character recognition threshold failures with excellent precision.

In comparison, the graph also reveals that there is a clear discrepancy between the combinations of the results between certain human participants and current simulation configurations, resulting in Poor, Slight, or Fair agreement. This pattern is distinctly visible across the entire data tracks for Participant 03 (values ranging from 0.059 to 0.155). Under the established interpretation brackets, these values indicate that the simulation settings within these blocks failed consistently to accurately capture the specific reading limits or threshold breaking points of these two individual participants. This doesn't particularly indicate flaws in the study or the simulation pipeline, instead, this variance showcases that the VI-OCR pipeline is highly responsive to unique profile calibrations rather than outputting generic, arbitrary image degradation. In these cases, the HShift and VShift properties we utilized for the VI-OCR simulations did not accurately simulate the visual impairments of these two participants. By experimenting with other HShift and VShift properties, and the "tuning" of the VI-OCR filtering part of the pipeline, we could potentially achieve results with a higher Kappa, showing more agreement.

Finally, an interesting anomaly can be observed regarding simulation configurations 05, 09, and 10. Across all four human participants, the calculated Kappa coefficients for VI-OCR-05, VI-OCR-09, and VI-OCR-10 are perfectly identical, resulting in static values of 0.028 for P01, 0.493 for P02, 0.155 for P03, and 0.567 for P07. Mathematically, this behavior indicates that the parameter variations introduced between these three simulations had practically no impact on the OCR engine's ability to recognize our filtered inputs. If we look closely at the values for HShift and VShift for each of these simulations: VI-OCR09 (H_SHIFT = 4.8, V_SHIFT = 0.065); VI-OCR10 (H_SHIFT = 6.2, V_SHIFT = 0.085); VI-OCR05 (H_SHIFT = 5.5, V_SHIFT = 0.075); These are indeed fairly similar. So, in summary, although the VI-OCR pipeline worked as expected, simulating these three profiles similarly, it should be further "tuned" to ensure it can better mimic the nuances between individuals with similar types of visual impairments.

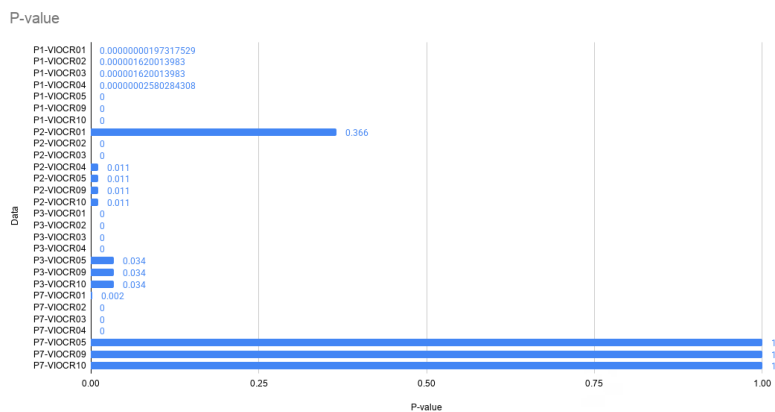


Fig.46. Comparison of P-Value results for each participant and VI-OCR simulation combination.

To better understand if the Kappa results we obtained were due to chance, it is important to look at the results from the P-value graph (Figure 46).

Most P01 related simulations produced extremely small p-values. These values are far below the significance threshold and indicate strong statistical evidence that the VI-OCR simulations did not accurately replicate the participant's observed performance. This suggests that P01 always recognizes the trigrams where the machine fails, or *vice versa*. The simulation may fail to capture im-

portant aspects of P01’s visual impairment, or maybe the participant-specific visual behavior is more complex than the simulations.

P02 shows mixed simulation performance, with P2-VIOCR01 showing a P-Value of 0.366, consequently being one of the strongest results in the graph. The VI-OCR simulation may accurately reproduce this participant’s behavior, or at minimum, the experiment could not detect a meaningful statistical difference. Other P02 simulations yielded either P-Values of 0 or 0.011. Although the differences for the simulations with a P-Value of 0.011 may not be extremely large, the simulations still differed detectably from participant performance. Interestingly, these values are not nearly as extreme as the P01 results, suggesting that the VI-OCR pipeline may have performed better for P02 overall than for P01, though still imperfectly.

Most P03 simulations show extremely small P-Values near zero, again indicating strong disagreement between the simulation and actual participant performance. P07 presents the most interesting pattern in the entire graph, some P-Values indicate again, strong disagreement, however several simulations produced extremely high P-Values of 1, which are the strongest in the data set suggesting that the participant and simulation distributions are extremely similar statistically, with essentially no detectable difference.

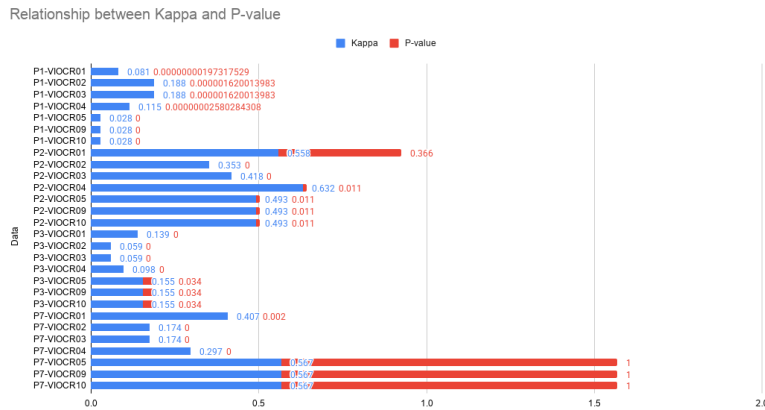


Fig.47. Relationship Between Kappa and P-Value.

As previously mentioned, the primary objective of this study was to determine to what extent the quantitative legibility metrics generated by the VI-OCR computational simulations reliably correlate with data obtained from human participants. More specifically, the study aimed to validate the central hypothesis that the VI-OCR pipeline can function as a scientifically grounded method for simulating impaired human vision in legibility testing scenarios. The relationship between the Kappa coefficients and the P-Values (Figure 47) is therefore particularly important, as it allows us to determine not only whether the simulations align with human behavior, but also whether such alignment is statistically meaningful.

Among all combinations analyzed, the pairing P2-VIOCR01 emerged as one of the strongest validation outcomes in the entire dataset. This combination produced a Kappa coefficient of $\kappa = 0.558$, indicating moderate-to-strong agreement, while simultaneously generating a non-significant p-value of 0.366. This combination is scientifically important because it demonstrates both practical agreement and statistical consistency between the participant and the simulation. In other words, the VI-OCR01 simulation was able to reproduce the recognition behavior of P02 without exhibiting statistically significant discrepancies.

An even more interesting result was observed in the pairing P2-VIOCR04, which produced the highest Kappa coefficient in the entire study with $\kappa = 0.632$, representing substantial agreement. This indicates that the simulation and the participant shared a highly similar recognition pattern, repeatedly succeeding and failing on similar visual recognition tasks. However, despite this strong agreement, the corresponding p-value was statistically significant at 0.011. Although the simulation clearly captures the general structure of Participant 2's impairment profile with considerable precision, the significant p-value indicates that subtle but measurable differences still exist between the simulation behavior and the participant's real visual performance. Scientifically, this suggests that the simulation is functioning as a high-fidelity approximation rather than as a perfect duplicate of the participant's visual impairment.

Several additional pairings involving P02, namely P2-VIOCR05, P2-VIOCR09, and P2-VIOCR10, also produced meaningful results. These simulations generated Kappa values around $\kappa = 0.493$ while maintaining statistically significant p-values of 0.011. These outcomes indicate moderate agreement between the simulation and the participant, suggesting that the VI-OCR pipeline is successfully reproducing many aspects of the participant's visual impairment profile. However, the statistically significant p-values demonstrate again that measurable differences still remain between the computational simulations and the participant's actual visual behavior.

In contrast, P01 consistently produced low Kappa coefficients ranging from $\kappa = 0.028$ to $\kappa = 0.188$, accompanied by extremely small p-values approaching zero. Statistically, this combination represents one of the weakest validation scenarios observed in the study. The low Kappa values indicate weak agreement between the participant and the simulation outputs, while the extremely small p-values confirm that the differences between the participant and the simulations are systematic rather than random. In practical terms, the VI-OCR configurations used in these conditions failed to accurately reproduce the recognition thresholds and reading limitations associated with P01's visual impairment profile.

A similar pattern was observed for P03, whose simulations generated low-to-fair Kappa coefficients alongside statistically significant p-values. This again suggests that the VI-OCR simulations partially captured the participant's visual behavior but failed to fully reproduce the specific impairment characteristics associated with this participant.

One of the most interesting observations in the entire dataset concerns simulation configurations for VI-OCR05, VI-OCR09, and VI-OCR10. Across all participants, these three simulations produced perfectly identical Kappa coefficients. Specifically, P01 consistently generated $\kappa = 0.028$, P02 produced $\kappa = 0.493$, P03 yielded $\kappa = 0.155$, and P07 produced $\kappa = 0.567$ across all three simulation configurations. This statistical behavior is highly informative because it indicates that the parameter variations introduced between these simulations had effectively no measurable impact on the OCR engine's recognition behavior. This, of course, corroborates what we were discussing before when examining the associated HShift and VShift values. These parameter values are sufficiently similar that they appear to fall within a stable operational range where the OCR engine responds almost identically.

P07 presents another particularly important case within the dataset. Several P07 simulation pairings, especially P7-VIOCR05, P7-VIOCR09, and P7-VIOCR10, produced high Kappa coefficients of $\kappa = 0.567$ alongside p-values equal to 1.0. This represents one of the strongest outcomes observed in the study. The moderate-to-strong Kappa values indicate meaningful agreement between the participant and the simulations, while the p-values of 1.0 suggest that no statistically detectable differences exist between the two behaviors.

7.1.2. Secondary Research Questions

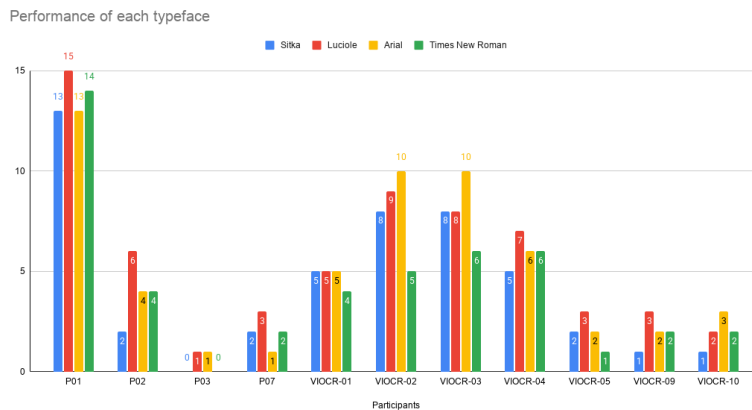


Fig. 48. Performance of each typeface.

The analysis of typeface behavior (Figure 48) provides additional insight into how the VI-OCR simulations performed in terms of typographic legibility when compared to human participants. When examining the overall performance of each typeface, clear differences emerge regarding how effectively certain letterforms withstand the imposed degradation conditions. Across both human participant testing and VI-OCR simulations, Luciole seems to have shown a strong overall performance, although, as previously mentioned, these differences cannot be interpreted as definitive due to the relatively small sample size used throughout the experiment, and should be explored in further research.

The consistency of these tendencies across multiple participant and simulation conditions suggests that the structural characteristics of each typeface may meaningfully influence recognition behavior under degraded viewing conditions. Luciole’s comparatively strong performance is particularly noteworthy given that the typeface was originally designed specifically to improve accessibility and readability for visually impaired readers, suggesting that its design principles may successfully preserve recognizable typographic structures even when subjected to substantial computational degradation. This tendency is especially visible in the participant-level data, where Luciole seems to have outperformed the remaining typefaces. Such behavior may suggest that design features commonly associated with accessibility-oriented typefaces, including exaggerated letter differentiation, increased openness of counters, and simplified structural forms, contribute positively to recognition stability under impaired visual conditions.



Fig.49. Comparison between MAE and MSD.

Interestingly, while its overall recognition rates remained comparatively high, the Mean Signed Deviation (MSD) analysis (Figure 49) revealed that the VI-OCR simulations tended to overestimate participant performance for Times New Roman. Specifically, Times New Roman produced the highest positive MSD value among all tested typefaces. This suggests that, under the applied degradation conditions, the OCR engine may process certain serif-related typographic structures less efficiently than human participants experiencing simulated visual impairments. In practical terms, when filtering input images, the VI-OCR pipeline may over-process typefaces with more structural details and subsequently under-perform when recognizing trigrams.

Arial, by contrast, produced one of the most interesting behaviors within the dataset. Although the typeface generated the highest Mean Absolute Error (MAE) value among all tested typefaces, it simultaneously produced a negative MSD value. The elevated MAE indicates that the VI-OCR pipeline's performance diverged more strongly from participant performance when compared to the remaining typefaces, suggesting lower predictive stability overall. However, the negative MSD reveals that the simulations systematically over-performed when recognizing the trigrams set in Arial. This may suggest that human readers were less able to compensate for their individual degradation effects when exposed to Arial than the VI-OCR pipeline itself. Sitka produced comparatively balanced results throughout the analysis. It generated one of the lowest MSD values, remaining very close to zero, which indicates that the VI-OCR simulations neither systematically overestimated nor underestimated participant performance to a substantial degree. From a simulation perspective, this is particularly interesting because it suggests that the VI-OCR pipeline processed Sitka in a manner more closely aligned with human participant behavior than was observed for several other typefaces. Although Sitka did not produce the absolute highest recognition performance overall, its low signed deviation may indicate a comparatively stable relationship between computational predictions and human visual recognition outcomes. The MAE analysis as a whole shows that all tested typefaces may have produced relatively similar average prediction errors, with values ranging approximately between 4.4 and 4.8.

Regarding the other secondary questions we set out to answer in the context of this study, particularly how feasible is the integration of the VI-OCR pipeline into the standard creative workflow of a type foundry, the learning curve for designers attempting to run these simulations independently, if there is demand within the type design industry for computational legibility tools, and the primary barriers to the adoption of such a protocol in professional practice, we did a few inquiries in order to get a better understanding of these questions. Those answers will be summarized alongside the other relevant information of this study in the discussion of findings sub-chapter.

Although we tried to reach as many relevant people as possible, not all of them answered or were available to answer our questions, which is understandable of course. With this in mind, I have to again thank Morisawa, Dave Crossland, Dan Rhatigan, Erik Van Blokland, Mary Meürer, Fabian Harb, Ricardo Santos, Octavio Pardo, Maíra Woloszyn, Gerry Leonidas, Erik Spiekermann, Aprígio Morgado, and James Edmondson for taking their time to answer my questions and share their insights.

7.2. Discussion of Findings

The strongest validation outcomes were observed primarily in configurations associated with P02 (Self-reported multi-pathological visual impairment with progressive visual deterioration) and P07 (Self-reported residual peripheral vision limited to the left eye). In particular, the pairing P2-VIOCR01 ($HShift = 4.1684$ and $VShift = 0.0575$) demonstrated moderate-to-strong agreement while simultaneously producing a non-significant p-value. From a scientific perspective, this represents one of the clearest indications that the VI-OCR pipeline can successfully approximate impaired visual recognition behavior.

The simulation was not only aligned with the participant's recognition patterns, but it also failed to produce statistically significant discrepancies when compared to the participant's performance. This result strongly supports the central hypothesis of the study and suggests that, under appropriate calibration conditions, the VI-OCR pipeline works at simulating VI vision.

Similarly, the configurations associated with P07 produced particularly compelling results. The combinations P7-VIOCR05, P7-VIOCR09, and P7-VIOCR10 generated moderate-to-strong Kappa coefficients together with p-values equal to 1.0. Statistically, this indicates that no detectable differences existed between the participant's recognition behavior and the outputs produced by the simulation. These outcomes suggest that the VI-OCR pipeline is capable of reproducing certain impairment profiles with considerable accuracy. More importantly, they indicate that the simulation pipeline is not simply generating generalized image degradation, but instead is capable of authentically reproducing participant-specific recognition tendencies under appropriate parameter configurations.

The results associated with P2-VIOCR04 (Self-reported multi-pathological visual impairment with progressive visual deterioration; $HShift = 3.0$ and $VShift = 0.035$) are also particularly important for understanding the behavior of the simulation pipeline. This configuration generated the highest Kappa coefficient observed in the entire study, demonstrating substantial agreement between the participant and the simulation. However, despite this strong behavioral alignment, the associated p-value remained statistically significant.

This combination reveals an important nuance regarding the interpretation of simulation fidelity. Although the simulation successfully reproduced the general structure of the participant's visual impairment, measurable discrepancies still remained between the participant's actual performance and the computational approximation. In practical terms, this suggests that the VI-OCR pipeline may already be capable of functioning as a high-fidelity approximation of impaired human vision, even if it has not yet achieved perfect participant-level replication.

At the same time, the weaker outcomes observed for Participants 01 and 03 reveal important limitations of the current simulation pipeline. Both participants consistently produced low Kappa coefficients accompanied by highly significant p-values, indicating weak agreement and systematic differences between the participant behavior and the simulation outputs.

These findings suggest that the filtering configurations used in the current study were unable to adequately reproduce the specific recognition thresholds associated with these participants.

However, this limitation should not necessarily be interpreted as evidence that the VI-OCR pipeline is fundamentally ineffective. Instead, the participant-dependent variability observed throughout the dataset strongly suggests that visual impairment simulations cannot yet be treated as a universal or one-size-fits-all problem. By what we gathered from our testing protocol, visual impairments are inherently heterogeneous, involving complex combinations of acuity loss, contrast sensitivity reduction, distortion, instability, and perceptual adaptation. Even participants with similar types of visual impairments may report substantial different difficulties, with the same participant reporting small daily variations of their vision that have to do with their environment, how well-rested they were, how they felt that day, how stressed they are and other potential variables. Consequently, it is reasonable that certain participants may require substantially different calibration strategies in order to achieve accurate simulation outcomes.

This participant-specific variability may, in fact, represent one of the most scientifically valuable findings of the study. Rather than artificially producing uniform agreement across all participants, the VI-OCR pipeline demonstrated a certain amount of sensitivity to calibration conditions and impairment profiles. Such behavior suggests that the pipeline is responding meaningfully to parameter modifications rather than generating arbitrary or random outputs. From a methodological perspective, this increases confidence in the authenticity of the obtained results and supports the idea that future improvements to the pipeline could further increase its predictive and simulation capabilities.

Another particularly relevant finding concerns the apparent stability observed across simulation configurations VI-OCR05, VI-OCR09, and VI-OCR10. Across all participants, these configurations produced perfectly identical Kappa coefficients despite using slightly different HShift and VShift values. This suggests that the parameter variations introduced between these simulations were insufficient to meaningfully alter the OCR engine's recognition behavior. In practical terms, the OCR system, which in our case was powered by Google Gemini's 2.5 flash, appears to operate within a relatively stable response region for these parameter ranges. This observation is important because it highlights both a strength and a limitation of the current simulation pipeline. On one hand, the consistency demonstrates that the VI-OCR pipeline behaves predictably under similar conditions. On the other hand, it also indicates that the current parameter ranges may not yet provide enough sensitivity to reproduce finer distinctions between visually similar impairment profiles.

The findings of this study also emphasize the importance of combining multiple statistical metrics when evaluating computational simulations of human perception. If only p-values had been examined, several simulations might have been prematurely classified as unsuccessful due to statistically significant differences. Conversely, relying exclusively on Cohen's Kappa could potentially overestimate simulation quality by focusing solely on agreement strength without accounting for measurable statistical discrepancies. The integration of both metrics therefore provides a more balanced and scientifically robust evaluation analysis, allowing the study to distinguish between simulations that merely exhibit superficial agreement and those that genuinely approximate participant behavior with statistical credibility.

Although they should be further analyzed and aren't set in stone because of the relatively small sample size, the findings obtained from the typeface analysis provide initial interesting insights into how specific typographic characteristics may influence both human and computational recognition performance

under degraded viewing conditions. One of the clearest tendencies observed throughout the analysis concerns the consistently strong performance of Luciole across both human participant testing and the VI-OCR simulations. The consistency of Luciole’s performance across multiple testing conditions suggests that accessibility-focused type design principles may genuinely improve recognition resilience under degraded visual environments.

While the VI-OCR pipeline primarily depends on structural consistency and feature extraction, human recognition is additionally influenced by perceptual adaptation, contextual interpretation, and familiarity with degraded forms. The Mean Absolute Error analysis further indicates that the VI-OCR pipeline maintained a relatively consistent level of predictive behavior across all tested typefaces. Despite observable differences in recognition tendencies and signed deviations, the average prediction errors remained relatively close throughout the dataset. This suggests that the simulation pipeline behaved coherently under the applied testing protocol rather than generating arbitrary or unstable outputs.

At the same time, the differences observed in the Mean Signed Deviation values demonstrate that the direction of the prediction discrepancies varied depending on the typeface being analyzed. The results may also indicate that the VI-OCR pipeline is sufficiently sensitive to detect behavioral differences between typefaces subjected to identical degradation conditions. This is particularly important because it demonstrates that the computational degradation protocol does not simply apply generic distortion, but instead interacts differently with distinct typographic structures. Such sensitivity strengthens the potential usefulness of the VI-OCR pipeline as an exploratory tool for evaluating typeface resilience, accessibility-oriented design strategies, and impaired-vision legibility performance. Nevertheless, as previously mentioned, these findings should be interpreted cautiously due to the limited scope of the present dataset. The observed tendencies should therefore be understood as preliminary indicators that need to be subjected to further research rather than definitive conclusions regarding the accessibility or readability superiority of specific typefaces.

The answers we got from our conducted inquiries provide interesting qualitative insight into how professionals and researchers within the fields of typography, accessibility, and type technology perceive the feasibility of integrating computational legibility tools such as the VI-OCR pipeline into real-world design workflows.

Across the interviews, a recurring pattern emerged: although there is considerable interest in the potential of automated accessibility validation systems, there is also strong caution regarding the limitations, risks, and methodological challenges associated with relying too heavily on computational evaluation alone. One of the clearest tendencies observed throughout the responses is that most participants do not view tools such as VI-OCR as replacements for the designer’s expertise or for human-centered testing, but rather as complementary instruments capable of expanding the designer’s perceptual awareness during the development process. Several interviewees emphasized that computational simulations could function as a type of “assistive layer” within the design workflow, helping designers visualize and identify accessibility problems that may otherwise remain invisible during conventional proofing processes. In this sense, the proposed role of VI-OCR shouldn’t be to automate design decisions, but instead to provide additional diagnostic feedback capable of informing typographic refinement and accessibility-oriented adjustments. The idea of integrating real-time legibility feedback directly into type design software was generally received positively, particularly when framed as a guidance system rather than a rigid evaluative authority.

At the same time, concerns were repeatedly raised regarding the possibility of designers becoming overly dependent on opaque scoring systems without fully understanding the underlying mechanisms responsible for the generated evaluations. Several participants warned that, without methodological transparency, computational scores risk becoming decorative metrics that encourage designers to optimize for machine recognition rather than for real human reading performance.

This tension between computational assistance and human judgment appeared consistently throughout the interviews. While some respondents viewed automated validation systems as a natural extension of existing quality assurance practices already present in contemporary font production pipelines, others expressed concern that accessibility metrics could unintentionally reduce typography to a simplistic optimization problem.

In particular, there was skepticism regarding the assumption that OCR recognition necessarily corresponds directly to human perceptual experience. Several participants emphasized that human reading involves contextual interpretation, perceptual adaptation, cultural familiarity, cognitive processing, and behavioral variability that are difficult to fully capture through computational models alone.

The inquires also revealed that many professionals already recognize accessibility and legibility as increasingly important concerns within large-scale typographic projects. Respondents working with major foundries, branding systems, and digital platforms noted that accessibility testing is already integrated into some contemporary workflows, particularly for products intended to be consumed by large and diverse audiences. At the same time, however, there appears to be no universally adopted protocol for evaluating typographic accessibility at the industry level.

This lack of standardized descriptors and methodologies was repeatedly identified as one of the primary barriers preventing the development of robust automated accessibility validation systems. Another important tendency emerging from the interviews concerns the perceived democratizing potential of open-source computational tools. Multiple respondents argued that systems such as VI-OCR could significantly reduce the financial and logistical barriers traditionally associated with accessibility testing and human clinical trials. Since large-scale legibility studies are often prohibitively expensive for independent designers, students, and small foundries, the availability of open-source simulation pipelines like VI-OCR could allow a broader range of practitioners to experiment with accessibility-oriented workflows and conduct preliminary testing without requiring access to institutional research infrastructure.

In this sense, the VI-OCR pipeline was frequently perceived not just as a technical tool, but as a mechanism capable of democratizing access to legibility research and accessibility experimentation. At the same time, it is clear that the adoption of these tools would likely depend heavily on workflow integration and usability.

Rather than expecting designers to operate standalone scientific software independently, many participants suggested that accessibility simulations would be more practical if embedded directly into existing design environments such as Glyphs, RoboFont, FontLab, or browser-based typography tools.

This perspective suggests that the primary learning curve associated with systems like VI-OCR may not necessarily stem from the complexity of the underlying algorithms themselves, but rather from how naturally the tools integrate into familiar design processes. In other words, if computational accessibility testing becomes embedded into existing proofing and quality assurance work-

flows, its adoption may become significantly more feasible for professional designers.

The flexibility provided by variable fonts was frequently described as particularly compatible with pathology-specific accessibility approaches, where typographic parameters could be adjusted according to the needs of individual users rather than optimized for a hypothetical “average” reader.

Despite the optimism surrounding these possibilities, ethical concerns were also consistently present throughout the responses. Several interviewees warned against the risk of computational simulations becoming substitutes for engagement with real visually impaired users. While many participants acknowledged that AI-based simulations could function as valuable empathy-building tools, they also stressed that accessibility research must remain grounded in human-centered testing and direct interaction with the communities being represented.

In this context, computational simulations were generally perceived as most valuable when used to supplement, reduce, or refine human testing procedures rather than eliminate them entirely. Overall, the questionnaire responses suggest that there is genuine interest within the type design, research and accessibility communities regarding the development of computational legibility validation tools such as the VI-OCR pipeline.

The interviews indicate that designers and researchers increasingly recognize the potential usefulness of accessibility-oriented simulations, particularly when integrated transparently into existing workflows and used as supportive rather than authoritative systems. However, the responses also demonstrate that significant barriers remain regarding methodological transparency, perceptual complexity, standardization, workflow integration, and ethical implementation. Consequently, the future adoption of computational legibility tools within professional typography practice will likely depend not only on improvements in technical accuracy, but also on the development of trustworthy methodologies, intuitive interfaces, educational integration, and continued human-centered validation practices.

The responses provided by the visually impaired participants add an important human-centered dimension to the findings of this study. Several participants confirmed that the difficulties experienced during the recognition tasks reflected challenges they encounter in everyday reading situations, particularly regarding letter size, contrast, and typeface selection in digital interfaces. Although some participants noted that the experimental setup could not fully replicate the complexity of real-world reading experiences, many still recognized parallels between the test tasks and the visual fatigue or recognition barriers they face daily.

This suggests that, despite the limitations of controlled experimental environments, the adopted test protocol utilizing a staircase procedure captured meaningful aspects of the interaction between VI participants and the shown stimulus. The participants also responded positively to the idea of researchers using AI-driven simulations to design more accessible typographic systems. Multiple responses emphasized that any initiative aimed at improving accessibility and reducing visual barriers was viewed favorably. At the same time, some participants highlighted an important limitation: while simulations and questionnaires may provide researchers with insight into visually impaired experiences, they cannot fully reproduce the lived reality of visual impairment itself. This observation reinforces the importance of maintaining human-centered validation practices alongside computational modeling approaches. Perhaps most significantly, the participants consistently associated accessibility-oriented typographic research with improvements in autonomy, usability,

and quality of life. Several responses emphasized the growing dependence on digital systems in everyday life and the frustration caused by inaccessible interfaces, poorly optimized typography, and applications that remain difficult to use even with magnification tools. In this context, the participants perceived research into accessible typography and computational legibility not just as a technical exercise, but as a potentially meaningful contribution toward digital inclusion and greater independence for visually impaired users.

7.3. Study Limitations, Constraints and Adopted Contingency Measures

Like all experimental research involving human perception, computational modeling, and accessibility evaluation, this study contains several methodological limitations and operational constraints that must be acknowledged when interpreting the obtained results. Although the findings provide meaningful insight into the potential of the VI-OCR pipeline as a computational legibility simulation framework, the present dissertation should primarily be understood as an exploratory and proof-of-concept investigation rather than as a definitive validation of universal impaired-vision simulation.

One of the primary limitations of the study concerns the relatively small participant sample utilized throughout the test protocol. Due to the practical constraints associated with recruiting visually impaired participants and conducting controlled legibility testing sessions, the sample size remained limited when compared to the scale typically required for large statistical generalization studies. Consequently, while the observed trends and correlations are meaningful within the scope of the experiment, the results should not be interpreted as universally representative of all visually impaired populations and conditions. The variability observed between participants further reinforces this limitation, since visual impairment itself is highly heterogeneous and influenced by numerous physiological, perceptual, and behavioral factors.

Related to this, the participant group may also contain characteristics that are exceptionally individualized and difficult to generalize computationally. Visual impairments are rarely experienced identically between individuals, even among users sharing similar clinical diagnoses. Factors such as adaptation strategies, reading habits, familiarity with digital interfaces, perceptual compensation mechanisms, fatigue, lighting conditions, and cognitive interpretation may substantially influence reading behavior in ways that are difficult to fully reproduce computationally. This may potentially partially explain certain unexpected results observed throughout the study, particularly the discrepancy in simulation fidelity across different participants.

Another important limitation concerns the scope of the number of typefaces utilized during testing. This dissertation does not aim to explore the entirety of the typographic design space, nor does it attempt to optimize every possible typeface or parameter combination. Instead, the study intentionally focused on a selected group of typefaces and specific generalized typographic characteristics in order to maintain methodological feasibility within the constraints of a master's dissertation. A comprehensive exploration of all typographic variables, accessibility-oriented features, and parametric combinations would require a substantially larger experimental protocol and considerably more extensive participant testing.

Similarly, the study does not attempt to validate the VI-OCR pipeline across all forms and severities of visual impairment, nor across all possible reading contexts. Consequently, the present findings should not be interpreted as evidence that the VI-OCR pipeline can accurately reproduce every visually impaired reading condition, especially in more complex real-world environments involving motion, glare, dynamic interfaces, environmental lighting variation, dense layouts, or prolonged reading fatigue.

Methodologically, the research was also intentionally limited to a single type of computational simulation pipeline and one primary human testing protocol. Our protocol using VI-OCR did not compare multiple OCR engines, we've opted to utilize Google Gemini as our OCR engine instead of other alternative AI recognition models (ChatGpt, Tesseract, etc). Likewise we have not compared multiple human legibility assessment methodologies.

This constraint was deliberately adopted in order to preserve methodological coherence and maintain a controlled comparison environment between the VI-OCR simulations and the participant testing procedure.

Expanding the study to include multiple simulation architectures and alternative human-testing methodologies would significantly increase both experimental complexity and the analysis scope beyond what was operationally feasible within the duration of the project and our own abilities. Although the VI-OCR pipeline demonstrated measurable correlations with participant behavior in several conditions, OCR-based recognition does not function identically to human reading.

Human perception involves contextual interpretation, perceptual adaptation, semantic prediction, prior familiarity, cognitive compensation, and behavioral flexibility that are difficult to fully replicate computationally. Consequently, even in situations where strong statistical agreement was observed, the simulations should not be interpreted as complete reproductions of human visual experience. Several operational constraints also affected the experimental process itself. Due to time limitations, resource availability, and the complexity associated with conducting accessibility-oriented user testing, it was not possible to perform large-scale longitudinal validation studies or extensive multi-session participant retesting, meaning that we had to extensively reduce the scope of our test protocol and study to match what's achievable in the time-frame we have available.

Additionally, certain variables could not be perfectly standardized across all testing conditions. Although contingency measures were implemented to minimize these inconsistencies, small variations in participant state and testing environment could potentially, although in a very small immeasurable way, still have influenced portions of the collected data.

Potential sampling limitations must also be acknowledged. Since participation relied partly on accessibility, availability, and voluntary recruitment, the study may contain sampling biases related to the demographic composition, digital literacy, or technological familiarity of the selected participants. Controlled testing procedures, repeated simulation conditions, cross-comparison between computational and participant data, and the combined use of multiple statistical evaluation metrics were implemented to improve analytical robustness. Furthermore, we intentionally maintained transparency regarding the scope and limitations of the VI-OCR pipeline, avoiding claims of universal applicability beyond the experimental conditions directly examined within the dissertation.

Last but not least, another limitation of our test protocol was the lack of medical experience to better evaluate participants' conditions. An attempt to reach out to specialists in optometry to provide help with our test protocol was made but unfortunately it was unsuccessful, meaning we had to rely instead on self-reporting of the participants visual impairments and a non-medical assessment of the Pelli-Robson and ETDRS test thresholds. The presence of these specialists could have been really interesting and helpful during the implementation of our test protocol and the subsequent analysis of acquired data.

IV.

Conclusion

8.

Summary of Results and Deliverables

This dissertation set out to study whether the VI-OCR computational simulation pipeline could function as a reliable and scientifically grounded alternative or complementary tool to traditional human-subject legibility testing for visually impaired readers. More specifically, the research sought to determine to what extent the quantitative legibility metrics generated by the VI-OCR simulations correlate with data obtained from human participants following a controlled trigram-based testing protocol derived from the methodologies proposed by Larson and Carter (2016).

Simultaneously, the dissertation aimed to evaluate the practical feasibility, professional applicability, and broader relevance of integrating computational simulations into contemporary type design workflows. The results obtained throughout the study suggest that the VI-OCR pipeline demonstrates meaningful potential as a computational approximation tool for impaired-vision legibility evaluation. The statistical analysis revealed several simulation-participant pairings that produced moderate-to-strong Cohen's Kappa coefficients alongside non-significant p-values, indicating situations where the computational simulations aligned closely with human participant recognition behavior.

These findings support the central hypothesis of the dissertation by demonstrating that, under specific conditions and parameter calibrations, the VI-OCR pipeline is capable of approximating aspects of impaired human visual recognition with measurable statistical credibility. At the same time, the study also revealed important limitations regarding the universality of the simulations. Certain participant profiles produced weak agreement coefficients and statistically significant discrepancies between the simulation outputs and human recognition performance.

These outcomes indicate that the VI-OCR pipeline does not yet function as a universally accurate representation of all visual impairments. It has shown a strong approximation to P02 (self-reported multi-pathological visual impairment with progressive visual deterioration) and P07, (self-reported residual peripheral vision limited to the left eye) human participants, P2-VI-OCR04, P7-VIOCR05, P7-VIOCR09 and P7-VIOCR10 with the parameters HShift = 3.0 and VShift = 0.035, HShift = 5.5 and VShift = 0.075, HShift = 4.8 and VShift = 0.065 and 6.2 and VShift = 0.085 respectively.

Hence, the results suggest that computational accessibility simulation remains highly dependent on participant-specific calibration and the careful tuning of degradation parameters. In this sense, the study demonstrates both the promise and the current limitations of VI-OCR.

Accessibility-oriented typefaces such as Luciole demonstrated comparatively strong recognition performance across both participant testing and simulation conditions, while other typefaces exhibited more complex relationships between human and computational recognition behavior. The results suggest that the VI-OCR pipeline is sufficiently sensitive to detect behavioral differences between typefaces subjected to identical degradation conditions, supporting its potential usefulness for exploratory accessibility-oriented typographic evaluation.

However, the study also demonstrated that OCR recognition does not always perfectly correspond to human perceptual behavior, particularly when more structurally detailed typefaces are involved. Beyond the statistical and experimental findings, the dissertation also produced several practical and methodological deliverables.

A complete test protocol integrating human-subject legibility testing, VI-

OCR simulations, OCR recognition analysis, and statistical comparison was successfully implemented and documented. The study additionally produced a comparative experimental dataset crossing simulation outputs with human participant performance across multiple typefaces and degradation configurations.

Furthermore, the research generated a structured methodological framework capable of supporting future experimentation involving computational legibility simulation and accessibility-oriented type design research.

The qualitative interviews and questionnaires conducted with professional type designers, researchers, accessibility specialists, and visually impaired participants further expanded the scope of the dissertation beyond the technical aspects. These responses revealed that there is genuine interest within the professional and academic communities regarding computational accessibility simulation tools, particularly when positioned as complementary systems capable of supporting rather than replacing human-centered accessibility testing. The interviews also highlighted several important concerns regarding transparency, methodological validity, ethical implementation, workflow integration, and the risk of reducing complex human perceptual experiences to overly simplified numerical metrics.

Collectively, the findings of this dissertation contribute to the growing intersection between typography, accessibility research, computational design methodologies, and AI-assisted evaluation systems, helping to further develop on-going international research. Although the present work does not claim to establish a universally validated accessibility simulation pipeline, it demonstrates that computational legibility models can produce statistically meaningful correlations with human recognition behavior under controlled experimental conditions. Consequently, the research supports the broader possibility that AI-assisted perceptual simulation may become an increasingly valuable component of future accessibility-oriented typographic workflows and design validation methodologies.

9.

Discussion of Results

As discussed in Section 7.2, the VI-OCR pipeline demonstrated varying levels of agreement across participants, suggesting that some visual impairment profiles are more readily approximated through computational simulation than others. One possible explanation is that impairments primarily affecting clarity, sharpness, or contrast align more closely with the image degradation mechanisms implemented within the pipeline. In contrast, visual impairments involving perceptual adaptation, visual instability, scanning behavior, or neurological factors may be less accurately represented through image processing alone. These findings reinforce the idea that computational accessibility simulations are highly dependent on participant-specific calibration and should not yet be considered universally applicable representations of visual impairment.

The results also highlight important differences between human and machine recognition. While the VI-OCR pipeline operates deterministically, human visual performance is inherently variable and influenced by factors such as fatigue, attention, adaptation, and environmental conditions. Additionally, the trigram-based methodology intentionally removed contextual linguistic support, preventing participants from relying on semantic prediction or sentence reconstruction strategies commonly used during everyday reading. Consequently, some of the discrepancies observed between simulated and human performance may reflect methodological differences between controlled recognition tasks and natural reading behavior.

Finally, the findings reinforce the distinction between OCR recognition and human perception. Human readers often compensate for degraded visual information through perceptual and cognitive mechanisms. As a result, computational simulations should currently be viewed as complementary accessibility evaluation tools rather than replacements for human-centered testing. Nevertheless, the statistically meaningful correlations observed throughout the study demonstrate that computational legibility simulations can provide valuable insight into impaired visual recognition and may contribute to the development of future accessibility-oriented type design workflows.

10.

Dissemination

The research developed throughout this dissertation has already begun to generate academic dissemination outcomes. Part of the work was presented publicly during IJUP 2026 (*Encontro de Investigação Jovem da Universidade do Porto*), where the study was introduced and discussed within an academic research context. The presentation provided an opportunity to communicate the objectives, methodology, preliminary findings, and broader implications of the VI-OCR pipeline to researchers and students from multiple disciplinary backgrounds, contributing to critical discussion surrounding accessibility-oriented computational design methodologies. In addition to IJUP's book of abstracts, the work associated with this dissertation will also be formally archived within the University of Porto's institutional repository, contributing to the long-term accessibility and visibility of the research. Beyond this, there will be an attempt made to get this research published in academic journals and other relevant mediums.

The potential of this study hopefully extends itself beyond purely academic contexts and may also contribute to conversations occurring within professional type design communities, accessibility initiatives, and digital product development environments. Finally, the exploratory nature of the VI-OCR pipeline, alongside the transparency regarding its limitations and methodological constraints, contributes to an open and constructive discussion concerning the future role of computational accessibility simulation within human-centered design methodologies. In this sense, the dissemination of the project is intended not only to communicate its findings, but also to encourage continued experimentation, interdisciplinary collaboration, and critical reflection regarding how accessibility-oriented AI systems may responsibly support future typographic research and inclusive design practices.

11.

Limitations,
Improvements
and Recom-
mendations
for Future
Work

Throughout this study we have mentioned that the testing of the VI-OCR pipeline was not easy and user friendly, we found many problems running either the original repository or the evaluatevisibility.com web version (Figure 50) taking months to find a way to correct these issues.

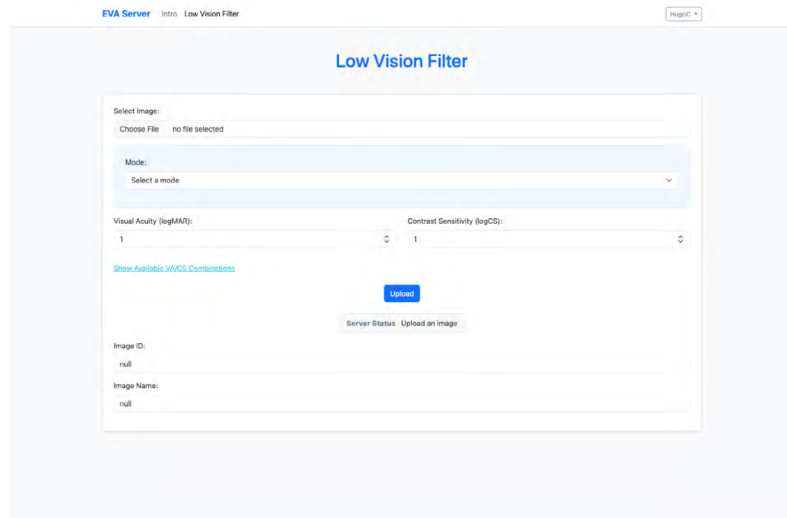


Fig.50. Website developed in the context of VI-OCR's original research (limited access). <http://evaluatevisibility.com/>

In order to advance we started further exploring the development of our previously mentioned web-based tool with new iterations (Figure 51), but found a lot of limitations whilst trying to achieve similar results to the initial paper presented by the researchers at Johns Hopkins University. Mostly, this had to do with the way it was applying the filtering to the input images. Also, because these tools were developed with the help of artificial intelligence and this falling way outside of our technical abilities, we couldn't in no way or shape affirm our code was flawless and doing what it was supposed to do with 100% certainty. To solve this there was an attempt to contact programmers to help with the development of the tool or to at least verify our code-base, but unfortunately that was not successful.

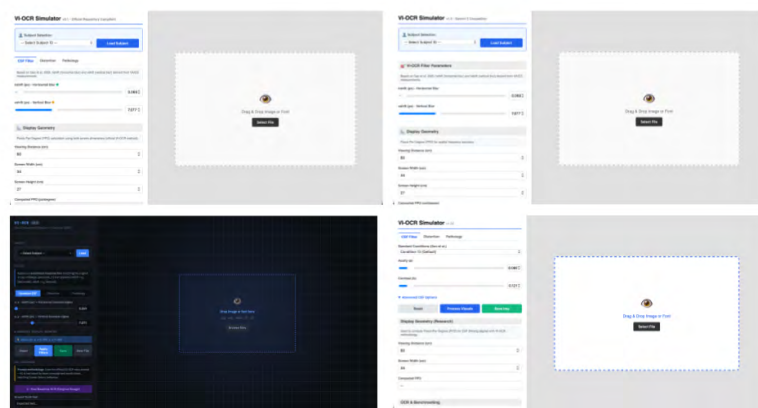


Fig.51. Iterations of our web-based VI-OCR implementation.

This is one of the biggest limitations of the VI-OCR pipeline, at the moment, you can't just access a website or install a tool to be able to run it. The evaluatevisibility.com web-tool, which we got limited access to, did not allow us to download images and do the OCR recognition part of the pipeline as it is in a very early stage of development.

When we first tried to install and run the VI-OCR pipeline from the repository, we quickly discovered that the provided code was incomplete and the instructions were often incorrect or misleading. The very first issue we encountered was that a core file called `inference_pipeline.py` was completely missing from the repository, despite being imported by the main script.

Without this file, nothing could run. So we had to build it from scratch, studying what the main script expected and writing our own implementation of the filtering and inference pipeline. The next major hurdle came from the Google Gemini integration. The repository instructed us to use `from google import genai` and to call the `gemini-2.0-flash` model. Neither of these worked. The import statement was outdated and failed immediately, and when we finally got past that, the model name itself had been deprecated and was no longer accessible to new users. We had to update the import to `import google.generativeai as genai` and switch to the current model `gemini-2.5-flash`. This was not mentioned anywhere in the documentation.

Similarly, the repository told us to import a function called `apply_vision_filter` from the utilities module, but when we looked inside that file, the function was actually named `add_filter`. This kind of naming inconsistency appeared repeatedly and required constant troubleshooting. Even getting the basic data files in place was problematic. The repository showed the CSV files sitting in a folder called `data/human`, but the script that processed them was hardcoded to look for them in the root `data/directory`. This meant that even after downloading the required CSV files from Google Drive and placing them exactly where the README indicated, the script still could not find them.

We had to manually copy the files into the location where the script actually expected them, a step that was documented nowhere. The annotation file that tells the pipeline which images to process presented a similar challenge. The batch processing script expected a very specific JSON structure with a top-level key called `images`, but nothing in the repository clearly stated this format or provided a working example.

We had to experiment and deduce the correct structure through trial and error, eventually landing on a format that included both `img_name` and `id` fields for each image. Beyond the code itself, the repository failed to list all the necessary dependencies. We followed their installation instructions but still encountered a `ModuleNotFoundError` for `cv2`, which is OpenCV. This library is essential for image processing within the filter functions, yet it was not mentioned anywhere in the requirements. We had to manually install `opencv` along with other missing packages like `retry` and `matplotlib` as errors appeared as we progressed.

The repository's example prompt asked the model to identify "English words" in the image. However, our test images contained random three-letter trigrams, not real words. So when we first ran the pipeline, Gemini would look at each image, see no actual English words, and return an empty list every single time. The pipeline appeared to be completely broken, but in reality it was doing exactly what we asked: looking for English words that did not exist. We had to rethink the prompt, rewriting it to explicitly ask for "sequences of three letters" or "trigrams" and to clarify that these were random letter combinations, not real words. We also added conservative instructions telling the model to only output trigrams it was absolutely certain about and to never guess when the text was blurry or unclear. This change allowed the pipeline to successfully work across all sixty input images.

Finally, the provided interactive script `BATCH_PROCESS.sh` turned out to be completely unusable for our custom dataset. It was hardcoded to look for specific files from the original research dataset, such as `data/totaltext/anno`.

json, and gave us no reasonable way to override these paths for our own test images.

Every time we tried to use it, it would crash looking for files that did not exist. We eventually abandoned it entirely and wrote a simple, focused Python script that gave us full control over every parameter: where to find the images, what filter settings to apply, which prompt to use, and where to save the filtered images and results. This custom script successfully processed all sixty images while saving both the trigram results and the filtered images for later analysis. In the end, the only parts of the original repository that worked as intended were the core filtering algorithm itself and the model factory pattern. Almost everything else was missing, outdated, incorrectly documented, or simply broken. We had to debug each error as it appeared, building the missing components ourselves, and adapting the tool to the specific reality of our trigram-based dataset. The concept behind VI-OCR is sound and powerful, but the implementation provided by the repository required extensive work to become functional.

So, in many ways we need to look at the conducted research of this study taking into consideration we utilized a modified version of the original VI-OCR pipeline present in the repository made available on Github, which means that the VI-OCR pipeline in its current state was frustrating to setup, needed extensive fixes and the results we produced on each of the 10 simulations we ran have to be further corroborated with future testing.

From a personal point of view I don't see this as something negative or positive, I see it as an accurate description of the current possibilities and found issues of the VI-OCR pipeline and possible paths moving forward. One thing is clear, in order to be implemented and useful, VI-OCR needs to be fixed and user friendly for type designers. A path forward seems to be the evaluatervisibility.com website we've previously mentioned that is being developed, if the full pipeline is implemented and if it produces similar results to the ones found in this study and further testing we can assume we would have a much user-friendly way of simulating visually impaired vision.

Other aspects of the present study could be improved or expanded in future research iterations in order to strengthen the robustness, generalizability, and practical applicability of the obtained findings.

One of the most significant improvements would involve increasing the size and diversity of the participant sample. A larger participant pool representing a broader range of visual impairments, ages, nationalities, reading habits, and technological familiarity would allow for more statistically reliable conclusions and reduce the influence of participant-specific variability observed throughout the current study. The study could also benefit from the inclusion of additional visual impairment profiles and more clinically differentiated participant groups. Since visual impairment manifests differently across individuals and pathologies, future research could explore more granular impairment categorization, including distinctions between acuity loss, contrast sensitivity reduction, central vision impairment, peripheral vision loss, and other pathology-specific visual behaviors.

Another important improvement would involve expanding the number of analyzed typefaces. This study examined only a limited selection of typefaces and typographic parameters.

Future studies could analyze a substantially wider range of serif, sans-serif, variable, accessibility-oriented, and experimental typefaces, alongside more extensive exploration of typographic variables such as weight, width, spacing, x-height, contrast, optical size, and character differentiation.

With this in mind, even more interesting questions can and should be posed, like how the performance of VI users that utilize other scripts (CJK, Devanagari, etc) would compare to VI-OCR simulations.

The computational pipeline itself could also be refined. Additional filtering methods, more sophisticated degradation models, and adaptive simulation parameters could improve the VI-OCR pipeline's ability to reproduce participant-specific visual behaviors. Similarly, future versions of the pipeline could incorporate machine learning approaches capable of dynamically calibrating simulations according to individual participant profiles rather than relying exclusively on predefined filtering parameters. The research design could additionally be strengthened through the comparison of multiple OCR engines and computational recognition systems.

Future studies could also incorporate multiple human legibility testing methodologies rather than relying on a single experimental protocol. Reading speed analysis, continuous-text comprehension tasks, eye-tracking, fatigue evaluation, contextual reading assessments, medical assessment of the participants' visual impairments and usability testing could provide a more comprehensive understanding of how computational simulations can better emulate VI individuals, relate to real-world reading experiences and how they can be implemented in the market.

Several environmental and operational variables could likewise be more tightly controlled in future experiments. Factors such as display calibration, ambient lighting conditions, participant fatigue, screen quality, viewing distance, and session duration may all influence recognition performance and could be standardized even more rigorously in larger experimental settings. Another important improvement concerns the duration and structure of participant testing. Longitudinal studies involving repeated testing sessions over extended periods could help determine whether participant adaptation, familiarity, or learning effects influence the relationship between human performance and computational simulation outcomes.

From a workflow perspective, the VI-OCR pipeline could also, as previously established, be made more accessible and practical for non-technical users. The current pipeline still requires a relatively technical setup and familiarity with simulation parameters. Future development could focus on creating more intuitive interfaces, integrated design-tool plugins, real-time feedback systems, and simplified usability workflows that would allow professional designers to run accessibility simulations more independently and efficiently.

Future research could also focus on establishing shared datasets, standardized impairment simulation procedures, and common statistical evaluation frameworks capable of supporting reproducible accessibility-oriented typographic research. Another promising avenue would involve evaluating the educational potential of accessibility simulation tools within typography and design pedagogy. Future studies could investigate whether exposure to computational impairment simulations influences how students approach accessibility-oriented design decisions, potentially helping to increase awareness regarding inclusive typography practices during professional training.

Finally, future research could further strengthen the human-centered dimension of the study by involving visually impaired participants more directly throughout the iterative development process of the computational pipeline and implementation methodologies. Greater participatory involvement could help ensure that future simulation systems remain aligned with lived visual experiences rather than relying predominantly on computational approximations of impairment behavior.

V.

End-Matter

1.

References

- Afkari-Fahandari, A., Shabaninia, E., Asadi-Zeydabadi, F., & Nezamabadi-Pour, H. (2025). A *Comprehensive Survey of Transformers in Text Recognition: Techniques, Challenges, and Future Directions*. *ACM Computing Surveys*. <https://doi.org/10.1145/3771273>
- Amado, P. (2022). *Responsive Vox: Extending the Vox-ATypI Typographic Classification System into a Four-Tier Cartesian Model for an Interactive Educational Application*. *Springer Series in Design and Innovation*, 581–593. https://doi.org/10.1007/978-3-031-20364-0_49
- Beier, S. (2012). *Reading letters: designing for legibility*. Bis.
- Beier, S., Berlow, S., Boucaud, E., Bylinskii, Z., Cai, T., Cohn, J., Crowley, K., Day, S. L., Dingler, T., Dobres, J., Healey, J., Jain, R., Jordan, M., Kerr, B., Li, Q., Miller, D. B., Nobles, S., Papoutsaki, A., Qian, J., & Rezvanian, T. (2022). *Readability Research: An Interdisciplinary Approach*. *Foundations and Trends® in Human–Computer Interaction*, 16(4), 214–324. <https://doi.org/10.1561/11000000089>
- Beier, S., & Larson, K. (2010). *Design improvements for frequently misrecognized letters*. *Information Design Journal*, 18(2). <https://doi.org/10.1075/idj.18.2.03bei>
- Bilak, P. (2025). *What font users actually want: findings from Fontstand 2025 survey*. *Fontstand News*. <https://fontstand.com/news/design-news/survey/>
- Braille Institute. (2023). *Atkinson Hyperlegible*. <https://www.brailleinstitute.org/freefont/>
- Brinson, S. (2019, March 27). *The Font that Helps You Remember*. Sam Brinson. <https://www.sambrinson.com/sans-forgetica/>
- Campbell, F. W., & Robson, J. G. (1968). *Application of Fourier analysis to the visibility of gratings*. *The Journal of Physiology*, 197(3), 551–566. <https://doi.org/10.1113/jphysiol.1968.sp008574>
- Casutt, R. (2022, July 21). *BR for You. Bionic Reading*. <https://bionic-reading.com>
- Cattell, J. M. (1886). *The Time Taken Up by Cerebral Operations*.
- Cheng, K. (2005). *Designing type*. Yale University Press.
- Cooreman, B., & Beier, S. (2024). *A Theory of Visual Attention Based Assessment of Font Style: How important is x-height for font legibility?*. Poster session presented at Society for the Scientific Study of Reading: Annual Conference, Copenhagen, Denmark.
- Crossland, D. (2008). *The Free Font Movement [Dissertation]*. https://davelab6.github.io/matd-dissertation/dave_crossland_matd2008_dissertation.pdf
- Ferreira, C. (2014). *Formas (In)visíveis: A influência da forma das serifas na legibilidade de textos longos*.
- Fontlab Ltd. (2026). *Manual - FontLab 8*. *Fontlab.com*. <https://help.fontlab.com/fontlab/8/manual/>

- Freitas, V. (2017). *Design português de tipos digitais: (1990–2010): da procura de sinais identitários à construção de uma plataforma de divulgação*. <http://hdl.handle.net/10773/19167>
- Dömötör, G. & Horváth, A. (2025). *Applicability of a computer retinal model for scale-dependent investigation of legibility problems*. *Heliyon*, 11(4), e42484–e42484. <https://doi.org/10.1016/j.heliyon.2025.e42484>
- Gao, Q., Manduchi, R., Ramulu, P., Legge, G., & Xiong, Y. (2025). *VI-OCR: “Visually Impaired” Optical Character Recognition Pipeline for Text Accessibility Assessment*. *Research Square*. <https://doi.org/10.21203/rs.3.rs-7032700/v1>
- Noordzij, G. (2005). *The stroke: theory of writing* (P. Enneson, Trans.). *Hyphen*. (Original work published 1985)
- Gill, E. (2013). *An essay on typography*. Penguin Books. (Original work published 1931)
- Glyphs GmbH. (2024). *Glyphs Handbook*. Glyphsapp.com. <https://handbook.glyphsapp.com>
- Google Fonts. (n.d.). *Quality Assurance*. Google Fonts Guide. <https://googlefonts.github.io/gf-guide/qa.html>
- Google Fonts. (2011). *Roboto*. Google Fonts. <https://fonts.google.com/specimen/Roboto/about>
- Grilli Type. (2026). *GT Mechanik*. <https://gt-mechanik.com>
- Holmes, K. (2018). *Mismatch: How Inclusion Shapes Design*. The MIT Press.
- Horváth, A., & Dömötör, G. (2017). *Computer model for font legibility measurement*. *Intelligent Decision Technologies*, 11(3), 335–344. <https://doi.org/10.3233/idt-170299>
- Hudson, J. (2018, April 4). *Introducing OpenType Variable Fonts*. VariableFonts.io. <https://medium.com/variable-fonts/https-medium-com-tiro-introducing-opentype-variable-fonts-12ba6cd2369>
- ISO. (2019). *Ergonomics. Accessible design. A method for estimating minimum legible font size for people at any age*. <https://doi.org/10.3403/30342431>
- Jones, P. R., & Ometto, G. (2018). *Degraded Reality: Using VR/AR to simulate visual impairments*. *City Research Online* (City University London). <https://doi.org/10.1109/var4good.2018.8576885>
- Karow, P. (1994). *Digital Typefaces*. Springer.
- Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). *The neural bases of spatial frequency processing during scene perception*. *Frontiers in Integrative Neuroscience*, 8. <https://doi.org/10.3389/fnint.2014.00037>
- Krosl, K., Elvezio, C., Luidolt, L. R., Hurbe, M., Karst, S., Feiner, S., & Wimmer, M. (2020). *CatARact: Simulating Cataracts in Augmented Reality*. *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. <https://doi.org/10.1109/ismar50242.2020.00098>

- Larson, K. (2004).** *The Science of Word Recognition; or how I learned to stop worrying and love the bouma*. <https://learn.microsoft.com/en-us/typography/develop/word-recognition>
- Larson, K., & Carter, M. (2016).** *Sitka: a collaboration between type design and science*. *WORLD SCIENTIFIC EBooks*, 37–53. https://doi.org/10.1142/9789814759540_0003
- Legge, G. E., & Bigelow, C. A. (2011).** *Does print size matter for reading? A review of findings from vision science and typography*. *Journal of Vision*, 11(5), 1–22.
- Leonidas, G. (2013, December 1).** *A few things*. [Blog post]. Retrieved from <https://leonidas.net/2013/12/01/a-few-things/>
- Leonidas, G. (2015).** *On minimum quality in typeface design*. Presented at *ATypI Barcelona 2014*. [Presentation slides]. Retrieved from <https://speakerdeck.com/gerryleonidas/on-minimum-quality-in-typeface-design>.
- Leonidas, G. (2015).** *On systems and quality in typeface design*. *5th Meeting of Typography – Book of Proceedings*. *5th Meeting of Typography, Barcelos, Portugal*. https://www.academia.edu/16275721/On_systems_and_quality_in_typeface_design
- Li, M., Tengchao Lv, Chen, J., Cui, L., Lu, Y., Dinei Florêncio, Cha, Z., Li, Z., & Wei, F. (2023).** *TrOCR: Transformer-Based Optical Character Recognition with Pre-trained Models*. *Proceedings of the ... AAAI Conference on Artificial Intelligence*, 37(11), 13094–13102. <https://doi.org/10.1609/aaai.v37i11.26538>
- Luckiesh, M., & Moss, F. K. (1939).** *The visibility and readability of printed matter*. *Journal of Applied Psychology*, 23(6), 645–659. <https://doi.org/10.1037/h0055273>
- Lund, O. (1999).** *Knowledge Construction in Typography: The case of legibility research and the legibility of sans serif typefaces*.
- Lupton, E. (2010).** *Thinking with type: a critical guide for designers, writers, editors, & students*. (2nd ed.). Princeton Architectural Press.
- McClelland, J. L., & Johnston, J. C. (1977).** *The role of familiar units in perception of words and nonwords*. *Perception & Psychophysics*, 22(3), 249–261. <https://doi.org/10.3758/bf03199687>
- Microsoft. (2022, March 28).** *TrueType hinting - Typography*. Microsoft.com. <https://learn.microsoft.com/en-us/typography/truetype/hinting>
- Microsoft. (2024, May 31).** *OpenType specification (OpenType 1.9.1) - Typography*. Microsoft.com. <https://learn.microsoft.com/en-us/typography/opentype/spec/>
- Napoleão, E. (2021).** *Identidades Tipográficas: Modelo de suporte ao projeto tipográfico para lugares*. <https://hdl.handle.net/10216/141626>
- Neudecker, C., Baierer, K., Gerber, M., Christian, C., Apostolos, A., & Stefan, P. (2021).** *A survey of OCR evaluation tools and metrics*. *The 6th International Workshop on Historical Document Imaging and Processing*. <https://doi.org/10.1145/3476887.3476888>

- O'Regan, K. (1979). *Moment to Moment Control of Eye Saccades as a Function of Textual Parameters in Reading. Processing of Visible Language.* https://doi.org/10.1007/978-1-4684-0994-9_3
- Pardo, O. (2025). *The Path to Tiphares (04): the knowledge behind Tiphares.* Pathfinders.rocks; Pathfinders. <https://pathfinders.rocks/blog/the-path-to-tiphares-04-the-knowledge-behind-tiphares/>
- Pelli, D. G., & Bex, P. (2013). *Measuring contrast sensitivity. Vision Research, 90, 10–14.* <https://doi.org/10.1016/j.visres.2013.04.015>
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). *Crowding is unlike ordinary masking: Distinguishing feature integration from detection. Journal of Vision, 4(12), 12.* <https://doi.org/10.1167/4.12.12>
- Pelli, D. G., & Tillman, K. A. (2008). *The uncrowded window of object recognition. Nature Neuroscience, 11(10), 1129–1135.* <https://doi.org/10.1038/nn.2187>
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). *Crowding and eccentricity determine reading rate. Journal of Vision, 7(2).* <https://doi.org/10.1167/7.2.20>
- Perea, M., & Gomez, P. (2012). *Subtle Increases in Interletter Spacing Facilitate the Encoding of Words during Normal Reading. PLoS ONE, 7(10), e47568.* <https://doi.org/10.1371/journal.pone.0047568>
- PeterCon. (2022). *OpenType overview - Typography.* Learn.microsoft.com. <https://learn.microsoft.com/en-us/typography/opentype/>
- Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading.* Prentice Hall.
- Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of reading (2nd ed.).* Psychology Press.
- Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). *Eye movements, the perceptual span, and reading speed. Psychonomic Bulletin & Review, 17(6), 834–839.* <https://doi.org/10.3758/pbr.17.6.834>
- Roethlein, B. (1912). *The relative legibility of different faces of printing types. American Journal of Psychology, 23(1), 1–34.*
- Sherman, N. (n.d.). *Variable fonts are here. Now what? – Fonts Knowledge.* Google Fonts. https://fonts.google.com/knowledge/using_variable_fonts_on_the_web/variable_fonts_are_here
- Shi, B., Bai, X., & Yao, C. (2017). *An End-to-End Trainable Neural Net work for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11), 2298–2304.* <https://doi.org/10.1109/tpami.2016.2646371>
- Smith, R. (2007). *An Overview of the Tesseract OCR Engine. Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2, 2.* <https://doi.org/10.1109/icdar.2007.4376991>

- Soloway, J. (1997, April 23). *Adobe and Microsoft Announce Open Type Font Specification for Developers*. Source. <https://news.microsoft.com/source/1997/04/23/adobe-systems-and-microsoft-deliver-open-type-font-specification/>
- Solso, R. L. (1994). *Cognition and the visual arts*. Mit Press.
- Sousa, D. (2025). *Considerations on Condensing Typefaces*. <https://hdl.handle.net/10216/171593>
- Spiekermann, E. (2012). *Helvetica sucks*. Spiekermann.com. <https://spiekermann.com/en/helvetica-sucks/>
- Tai, Y. C., Sheedy, J., & Hayes, J. (2006). *Effect of letter spacing on legibility, eye movements, and reading speed*. *Journal of Vision*, 6(6), 994.
- Thompson, W. B., Legge, G. E., Kersten, D. J., Shakespeare, R. A., & Lei, Q. (2017). *Simulating visibility under reduced acuity and contrast sensitivity*. *Journal of the Optical Society of America A*, 34(4), 583. <https://doi.org/10.1364/josaa.34.000583>
- Tinker, M. A., & Paterson, D. G. (1932). *Studies of typographical factors influencing speed of reading, X: Style of type face*. *Journal of Applied Psychology*, 16(6), 605–613.
- Tracy, W. (1986). *Letters of Credit: A View of Type Design*. David R. Godine, Publisher. (Original work published 1968)
- Unger, G. (2007). *While you're reading*. Mark Batty Publisher.
- Unger, G. (2018). *Theory of type design*. Nai010 publishers.
- Velázquez, R., Sánchez, C. N., & Pissaloux, E. E. (2016). *Visual Impairment Simulator Based on the Hadamard Product*. *Electronic Notes in Theoretical Computer Science*, 329, 169–179. <https://doi.org/10.1016/j.entcs.2016.12.010>
- Wang, H., Pan, C., Guo, X., Ji, C., & Deng, K. (2021). *From object detection to text detection and recognition: A brief evolution history of optical character recognition*. *WIREs Computational Statistics*, 13(5). <https://doi.org/10.1002/wics.1547>
- Warde, B. (1955). *The Crystal Goblet, or Printing Should Be Invisible*. *Typo-L*. <http://gmunch.home.pipeline.com/typo-L/misc/ward.htm>
- Whitney, D., & Levi, D. M. (2011). *Visual crowding: a fundamental limit on conscious perception and object recognition*. *Trends in Cognitive Sciences*, 15(4), 160–168. <https://doi.org/10.1016/j.tics.2011.02.005>
- Woloszyn, M. (2022). *Variable fontwork: um framework para o processo de design de fontes variáveis*. <https://repositorio.ufsc.br/handle/123456789/240948>
- World Health Organization. (2019). *World report on vision*. In *World Health Organization* (pp. 1–180). <https://www.who.int/publications/item/9789241516570>
- Xiong, Y.-Z., Lei, Q., Calabrèse, A., & Legge, G. E. (2021). *Simulating Visibility and Reading Performance in Low Vision*. *Frontiers in Neuroscience*, 15. <https://doi.org/10.3389/fnins.2021.671121>

Ye, Q., & Doermann, D. (2015). *Text Detection and Recognition in Imagery: A Survey.* *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(7), 1480–1500. <https://doi.org/10.1109/tpami.2014.2366765>

2.

Appendices

Appendix A - Contingency Plan

Although in the case of this study there was not a need to move forward with the contingency plan, one was, as mentioned before, developed:

If we can't move past the "Pre-testing" phase, the dissertation will change its focus towards analyzing the character confusion matrices derived from the Larson testing protocol to identify and classify the typographic anatomical features (e.g., aperture, counter, terminal shape) that are the primary determinants of legibility failure for readers with diverse visual impairments.

Research Question

Which specific typographic anatomical features are most strongly implicated in character recognition errors and confusion patterns among visually impaired readers, and how can these empirical findings be synthesized into a systematic, evidence-based design model for accessible typeface development?

Hypothesis

Character confusion errors (e.g., 'c'/'e', 'a'/'o') will be overwhelmingly concentrated in characters with small, enclosed, or ambiguous anatomical features (specifically small apertures and high congestion points). The analysis will demonstrate that the relationship between a character's internal white space (counter) and its surrounding black shape (stroke) is the most critical variable for VI legibility, leading to the creation of a new, quantifiable design checklist.

Appendix B - Supplementary Material - Terminology

Typeface: The abstract, master design of the set of characters. It encompasses the collective stylistic attributes (e.g., serif, sans-serif, weight, contrast) that define the visual appearance of the script (Lupton, 2004).

Font: The operational mechanism or file that contains the typeface design. In digital systems, the font file (e.g., OpenType or TrueType) is the software used by operating systems and applications to render the characters.

Glyph: An individual character shape or element within a typeface. For instance, the letter 'a' may have multiple glyphs (e.g., a standard 'a' and a small-caps 'a'). The study of letter recognition, particularly in methodologies like the one utilized by Larson and Carter (2016) focuses specifically on the legibility of individual glyphs.

The distinction between OpenType (OTF) and TrueType (TTF) is fundamental to understanding the digital nature of fonts. The older TrueType (.ttf) format was developed in the late 1980s by Apple and Microsoft, utilizing quadratic Bézier curves to define character outlines, a technology optimized primarily for efficient screen rendering and cross-platform compatibility (Karow, 1994). However, the subsequent OpenType (.otf) format, a joint development by Adobe and Microsoft in the late 1990s (Soloway, 1997), is a super-set that represents the current industry standard. OpenType's key advantage lies in its flexibility, as its outlines can use either the same quadratic curves as TrueType or the more complex cubic Bézier curves (known as PostScript/CFF outlines), which offer robust control for high-resolution printing. Crucially, OpenType files are designed to contain vast amounts of advanced typographic data, allowing for sophisticated features, such as automatic contextual substitutions, discretionary ligatures, and advanced handling of complex scripts that are vital for modern, highly-engineered typefaces, whereas TrueType supports a more limited range of these features.

Therefore, OpenType is the dominant format for professional type design due to its superior capacity for integrating complex rules and design features into a single, scalable file. The evolution of these formats culminated in Variable Fonts, introduced as an extension of the OpenType specification (Hudson, 2018). A Variable Font is a single font file that contains a complete, continuous range of design variations (or 'instances') within it, allowing for fluid interpolation between multiple defined master styles. Instead of requiring separate files for different weights (e.g., Thin, Regular, Bold) or widths, a Variable Font uses design axes (such as Weight, Width, and Slant) to allow users or applications to select any point along these ranges (Sherman, n.d.).

For research concerning legibility, this technology is transformative, as it enables the systematic and fine-grained testing of hundreds of incremental design changes (e.g., testing stroke thickness at 10 intervals instead of just 3 preset weights) using the same foundational file, thereby generating far more detailed data for optical recognition computational models.

Kerning: The selective adjustment of space between specific pairs of characters to achieve visually consistent spacing and avoid problematic visual gaps.

Tracking (or Letter-spacing): The uniform adjustment of space across a range of characters or a block of text.

Set Width: The horizontal space occupied by a character, encompassing the width of the character's body and its intrinsic side bearings (the space immediately to the left and right of the character outline). Set width is a core component of overall text density.

x-height: The height of the lowercase letters (specifically the letter 'x') relative to the baseline. A larger x-height generally increases the visual “openness” of a typeface, making characters appear larger at smaller point sizes, which significantly influences legibility in low-resolution or low-vision scenarios (Beier, 2012a; Cooreman & Beier, 2024).

Counter (or counterform): The partially or fully enclosed white space within a character (e.g., the “hole” in an 'o' or 'p'). The size and shape of the counter are vital for character recognition; if the counter is too small, it may “fill in” at low resolutions or under visual degradation, leading to misidentification (Beier & Larson, 2010).

Stroke Weight (Blackness): The thickness of the strokes that form the character outlines. This is the primary driver of a font's “blackness” or visual weight. In computational models, this is often measured as the ratio of ink-to-whitespace.

Appendix B - Supplementary Material - Typographic Anatomy and Legibility

The analysis of legibility, particularly through computational means, demands a granular understanding of the letterform's anatomical structure. These components represent the physical parameters that directly influence a character's distinctiveness and are thus the primary inputs for any computational model attempting to simulate the complexities of human reading and visual impairment. As articulated by Lupton (2004), type anatomy is best understood through two essential categories: the horizontal metrics, which define the vertical space and relationships of the text block, and the structural elements, which define the unique geometry of the individual letterform.

The spatial relationships are organized around several key imaginary lines that guide type-setting. The baseline is the invisible line upon which all capital letters and the main body of the lowercase letters rest. The x-height is arguably the most critical dimension for legibility, as it refers to the height of the main body of a lowercase letter, specifically the height of the 'x' glyph. This measurement is paramount because a larger x-height increases the amount of visual area dedicated to the core character recognition zone, thereby often enhancing legibility, especially at small sizes or low resolution (Cooreman, B., & Beier, S., 2024). Extending from the x-height are the ascender and descender lines. Ascenders are the parts of lowercase letters (such as 'h' and 'l') that rise above the x-height, while descenders are the parts (such as 'p' and 'g') that drop below the baseline. These vertical extensions increase the differentiation between characters and words, a key factor in recognition and visual processing.

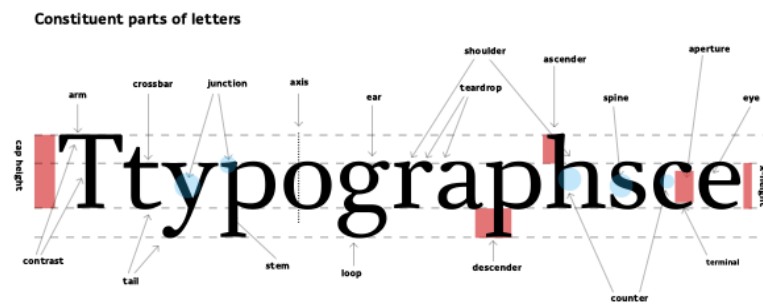


Fig.52. (Beier, 2012). *Constituent parts of letters [Diagram]*. In BIS Publishers (Ed.), *Reading Letters designing for legibility*.

Beyond these spatial metrics (vertical + horizontal), the letterform is defined by its internal and external structural elements. To explain the form and the stroke contrast inherent in any typeface, it is essential to reference Gerrit Noordzij's *The Stroke* (1985/2005), which theorizes that all typographic forms derive from the physical act of writing. Noordzij points out that variations in stroke thickness (contrast) are determined by two fundamental mechanisms: translation and expansion. The stem is the primary vertical or main diagonal stroke of a letter. A type design based on translation, analogous to a broad-nib pen held at a fixed angle, produces contrast where thick and thin strokes alternate based on the direction of the pen's movement, resulting in an angled axis of stress. Contrarily, a design based on expansion, analogous to the variable pressure of a pointed brush or pencil, creates contrast where stroke thickness is determined by pressure alone, typically yielding a vertical or near-vertical axis of stress (Gerrit Noordzij, 1985/2005). These principles govern the geometric foundation of the typeface.

With this in mind, the structural elements most central to modern legibility research are the negative spaces. The bowl is the rounded curve that encloses

the negative space, while the counter is the fully or partially enclosed negative space within a letter (e.g., the hole in an 'o' or an 'e'), and the aperture is the partially enclosed negative space that is open (e.g., the opening in a 'c' or 'e'). Research by Sofie Beier (2012) emphasizes that the clarity of these negative spaces, particularly the distinctiveness of the character shapes they create, is paramount to rapid character recognition. Beier's work shifts the focus from purely aesthetic or historical considerations to the psychological and perceptual factors, specifically highlighting how minimal geometric differences, such as the subtle variation between easily confused letters like 'a' and 'o', are vital for preventing visual confusion and improving reading speed.

This emphasis on distinctiveness and the geometry of these features ensures that the systematic measurement of character anatomy is directly aligned with the objective of maximizing legibility. Finally, the serif is the small finishing stroke added to the end of a main stem or arm, a feature that distinguishes serif typefaces and subtly influences the visual flow and weight of the text. This integrated structural vocabulary, underpinned by the theoretical basis of the stroke and the empirical findings on perceptual distinctiveness, allows for the precise deconstruction and systematic measurement of design variables.



Fig.53. (Beier, 2012). *Scientific Results [Diagrams]. In BIS Publishers (Ed.), Reading Letters designing for legibility.*

Appendix B - Supplementary Material - Distribution/Publication

Recently, a lot of foundries and marketplaces have started implementing subscription models allowing designers to test and “buy” typefaces in a more accessible way. Alternative models such as the one implemented by Adobe Fonts are also interesting to discuss, as they provide selected typefaces for “free” to sync, paying foundries per sync made on Adobe Fonts. Beyond these giants, alternative platforms such as Future Fonts have pioneered the “In-Progress” model, reflecting a shift toward “Living Fonts”. By releasing typefaces in iterations (v0.1, v0.2), designers can gather information on how a font performs in real-world environments, and get paid to continue the development of their typeface, mirroring the Early Access model found in high-level software development.

A significant philosophical departure from proprietary models is the Libre/Open-Source movement, which has been analyzed extensively by people like Dave Crossland (2008). Crossland argues that the SIL Open Font License (OFL) acts as a fundamental driver for typographic democratization and technical standard-setting. Unlike proprietary fonts governed by restrictive contracts, open-source fonts are treated as “software for the commons”.

This model has very interesting peculiarities, such as ensuring collaborative longevity. If, for example, a foundry ceases its operations, the source files remain available for the community to maintain, expand, or optimize. This collaborative approach allows for something unique in the industry, allowing anyone to contribute towards the development of a particular typeface.

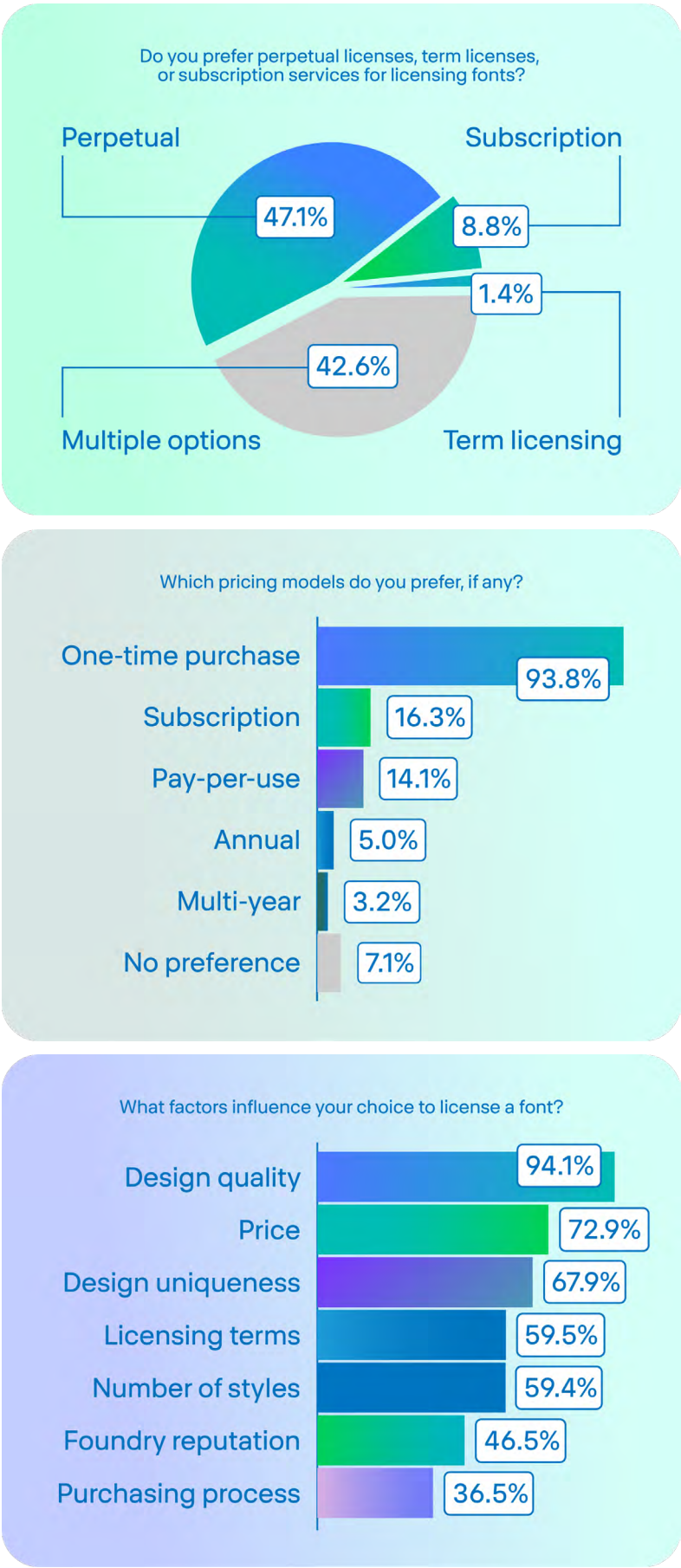
This can be particularly relevant for the development of minority scripts, where the open distribution allows native speakers and technical specialists to contribute to the character set, ensuring typefaces remain a resilient tool for global communication. Platforms such as Google Fonts have proved that this model works showing that their typefaces are mass adopted in millions of websites, an example of this is how, as of May 2026, the Google Fonts API served Roboto 62.2 billion times, and how Roboto is featured in more than 430 million websites worldwide (Google Fonts, 2011).

The mass adoption and funding of the Google Fonts library has translated into type designers contributing for this open-source project which has also made technical benchmarks, such as the previously mentioned audits made with, Fontbakery, “standards”, ensuring that fonts meet a high baseline of interoperability and performance.

The End User License Agreement (EULA) serves as the legal vessel through which a typeface exists as a digital product. It is a functional constraint that dictates the font’s technical reach, effectively acting as a “form of design” in itself. The EULA must balance the protection of the designer’s intellectual property with the user’s need for legal and technical certainty.

Modern licensing has evolved into a highly granular system, moving beyond the traditional “Desktop” license to include specific tiers for App embedding, Web server deployment, and Digital Broadcasting. This legal complexity is “hard-coded” into the font file’s Table ‘name’ and OS/2 table (fsType), which tell operating systems whether a font can be embedded or previewed. Mismanaging these fields during publication can result in “licensing conflicts”, for example, a font failing to render in a PDF or a browser, illustrating how the legal framework directly impacts the technical performance and perceived reliability of the typeface. This topic is subject to a lot of debate and endless discussion in the community and it was one of the “hot-topics” at the 2026 Fontstand Conference as Fontstand proposed an initial version of their new

licensing approach and showcased the findings of their Fontstand 2025 Survey (Figure 54).



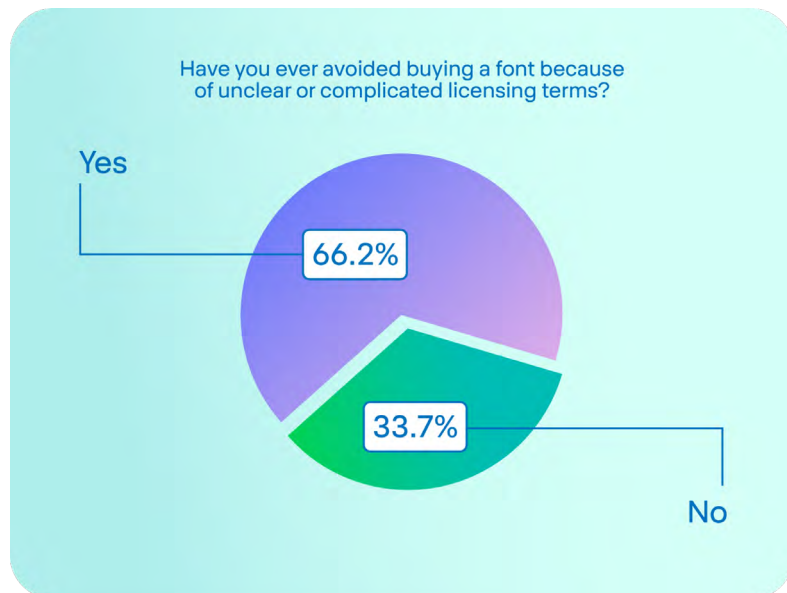


Fig.54. Findings from the 2025 Fontstand Survey. Source: Bilak (2025). <https://fontstand.com/news/design-news/survey/>

The publication stage is supported by a multi-layered communication strategy that translates technical engineering into perceived value. The traditional static specimen has been replaced by the interactive technical specimen or microsite (Figure 55), which serves as both a way to showcase the font’s underlying logic and also a marketing material. The specimen is an essential documentation of the designer’s intent, providing the end-user with the data required to utilize Open Type features, Variable axes, and Optical Sizes correctly.

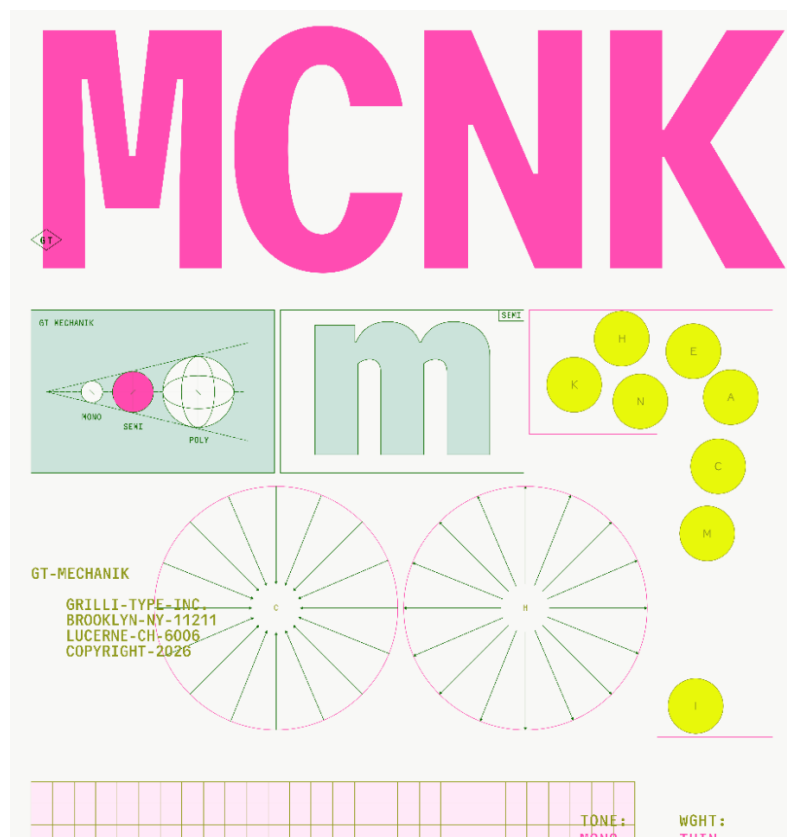


Fig.55. GT Mechanik Microsite. Source: Grilli Type (2026). <https://gt-mechanik.com/>

Beyond this type of marketing material, as previously mentioned, other ways of communication a new release may include social media marketing, blog posts, paid advertisements, forums and other alternative marketing methods.

Through my conversation with Dan Rhatigan, the topics discussed above were corroborated with a few extra valuable insights to the mix. As discussed above, the industry of type design offers a lot of different opportunities for the development and distribution of typefaces. According to Dan, marketing of a foundry can be simplified to the basics - social media posts, a well-built website with good SEO and a story for the typefaces you create, especially if you are starting out and don't have the funds to do paid ads, for example. Customers are interested in the narratives behind the foundries and the products they acquire. Dan also mentioned to me there are several paths, as discussed previously, that a type designer can take to promote and distribute their own work, noting that the decision on how to do this is something extremely personal that takes into account your own individual beliefs. Regarding his own type foundry, Bijou Type, Dan mentioned that for years he had concepts "in the shelf" of typefaces he wanted to create and release, it was only after leaving Monotype and joining Type Network that he finally kickstarted this project thanks to a distribution deal he made which also happened to include a partnership with Adobe Fonts. Regarding Adobe Fonts, Dan's perspective was very interesting as this business model can be very good for the promotion of foundries whilst generating income. Regarding the fact that you're essentially giving away your "free" typefaces, Dan mentioned that he looks at it from the perspective that these are his trial fonts for Bijou Type as the licensing for Adobe Fonts has a few key nuances that don't cover every licensing need. The Google Fonts model also offers a lot of visibility of your typefaces, in this case Dan referred me to Astloch and Copse, two typefaces he released in the early days of Google Fonts as two good examples of the type of visibility you get with both of them being served through Google's API a combined number of more than 6 million times over the last week (16th of may 2026).

Regarding licensing, from the conversations I had at Fontstand and my chat with Dan this can obviously be complicated, this was even one of the key discussion points in the conference. The important thing seems to be to make clear what the license covers and what it doesn't cover. If you want to distribute yourself, through your own website, this means that you have to take care of these nuances yourself, whereas if you opt to distribute through a marketplace or through an existing foundry these are concerns that you do not have to take into much consideration. At the end of the day, like Dan mentioned previously, this is dependent on what you want to do as a type designer and your own individual beliefs. You can and should take into consideration all the opportunities in front of you and specifically try to understand how you fit into all of this incredible world, perhaps you're not the most business-savvy individual, or you just want to draw letters, then marketplaces make a lot of sense for you. Perhaps you want full control over every single aspect of the customer experience, then, in this case, building a custom website and distributing by yourself may be the better option. Whatever you do it seems like the true important thing is to build good, interesting products that, as Dan advised me, tell a story and exist for some reason.

Appendix C - Stimuli (Original and Filtered from VI-OCR01 to VI-OCR10)

All the images used to create the dataset from the VI-OCR pipeline are available in this google drive folder: <https://drive.google.com/drive/folders/1O0y-gO4nzjNUKm6-xblw46SLgFSTm52c6?usp=sharing>

In this folder you will find the original images used (trigrams) and the images that were filtered with our “fixed” version of the VI-OCR pipeline presented, from VI-OCR01 to VI-OCR10, maintaining the same values for distance to the monitor, PPI and resolution, varying only the following HShift and VShift values:

VI-OCR01 - HShift = 4.1684 and VShift = 0.0575;
VI-OCR02 - Hshift = 1.0 and VShift = 0.008;
VI-OCR03 - HShift = 2.0 and VShift = 0.02;
VI-OCR04 - HShift = 3.0 and VShift = 0.035;
VI-OCR05 - HShift = 5.5 and VShift = 0.075;
VI-OCR06 - HShift = 7.0 and VShift = 0.09 (Removed because simulated vision was below the defined threshold);
VI-OCR07 - HShift = 8.5 and VShift = 0.11 (Removed because simulated vision was below the defined threshold);
VI-OCR08 - HShift = 10.0 and VShift = 0.125 (Removed because simulated vision was below the defined threshold);
VI-OCR09 - HShift = 4.8 and VShift = 0.065;
VI-OCR10 HShift = 6.2 and VShift = 0.085.

In this folder you will also find the resulting .json files from running the VI-OCR pipeline.



CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)

Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 16 de Abril de 2026

Assinatura do Participante: _____

Assinatura do Investigador: _____

CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)

Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

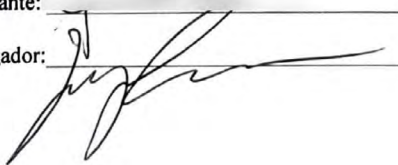
4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 16 de Abril de 2026 |

Assinatura do Participante: _____

Assinatura do Investigador: _____



CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)

Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 16 de April de 2026

Assinatura do Participante: _____

Assinatura do Investigador: _____

CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)

Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 16 de Abril de 2026

Assinatura do Participante: _____

Assinatura do Investigador: _____

CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)

Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

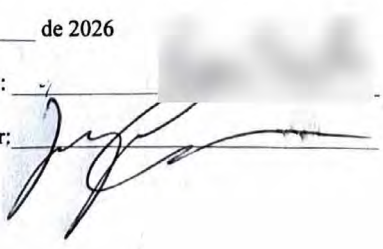
4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 23 de Abri de 2026

Assinatura do Participante: _____

Assinatura do Investigador: _____



CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)

Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 23 de Abril de 2026

Assinatura do Participante: _____

Assinatura do Investigador: _____

CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO

Projeto: Modelos de Simulação Computacional em Investigação da Legibilidade de Tipos de Letra - Avaliação da Fiabilidade como uma Alternativa a Testes com Participantes com Deficiência Visual

Investigador: Hugo Carvalho (Mestrado em Design Gráfico e Projetos Editoriais, FBAUP)

Orientador: Professor Dr. Pedro Amado (FBAUP)

1. Natureza e Objetivo do Estudo

Este estudo visa validar uma ferramenta digital (VI-OCR) que simula dificuldades visuais para testar a legibilidade de tipos de letra. O objetivo é comparar o desempenho de leitores com baixa visão com os resultados gerados pela simulação digital.

2. Procedimentos

Ao longo do protocolo de teste, irei realizar as seguintes tarefas:

- Responder a breves perguntas sobre a minha condição visual.
- Realizar testes básicos de acuidade e contraste (tabela ETDRS/Pelli-Robson).
- Identificar letras (trigramas) apresentadas num monitor a 150 cm de distância.
- Partilhar a minha opinião sobre a experiência num breve questionário final.

3. Voluntariedade e Confidencialidade:

- A minha participação é totalmente **voluntária** e não remunerada. Posso desistir voluntariamente a qualquer momento sem qualquer penalização.
- Caso me sinta desconfortável, posso pedir a qualquer momento, sem penalização, para o teste ser parado.
- Os dados recolhidos são **anónimos**. Não será guardado o meu nome ou qualquer dado identificável nos resultados finais.
- Os resultados serão utilizados exclusivamente para fins de investigação académica e publicações científicas relacionadas.

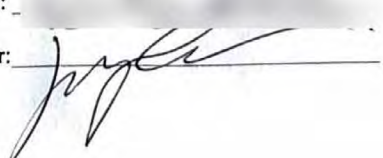
4. Riscos e Benefícios

Não existem riscos físicos e psicológicos associados. O benefício direto é contribuir para o desenvolvimento de ferramentas que ajudem designers a criar fontes mais acessíveis para pessoas com deficiência visual. Ao assinar, declaro que fui informado(a) sobre os objetivos do estudo, tive oportunidade de colocar questões previamente e durante a condução do mesmo e aceito participar voluntariamente. Ao abrigo do RGPD os participantes têm o direito de solicitar o acesso aos dados recolhidos.

Porto, 23 de Abri de 2026

Assinatura do Participante: _____

Assinatura do Investigador: _____



Appendix E - Test Protocol Data

Similarly to the Stimuli you can access the google sheets with all the data and its calculations here:

https://docs.google.com/spreadsheets/d/1rqzw4-UoYypjRNSSh1pAk4oZX-4GRK_XD3YGMNOzZTcw/edit?usp=sharing

A quick disclaimer, the Pre-test and the Screening tabs have notes written in Portuguese, the content of these notes was translated and previously mentioned when overviewing each participant, so please refer to that chapter.

All the calculations are in the Analytics tab.

Appendix F - Questionnaires/Interviews

Morisawa

Question: When developing a Latin typeface, designers often begin by defining structural parameters such as baseline, x-height, and proportions. For CJK, Indic, or other complex scripts or minority languages, compared to Latin scripts, what is the most important initial structural decision or visual constraint that must be established before detailed drawing begins?

Answer: CJK scripts have complex structures, including combinations of kanji and kana. Factors such as the glyph size relative to the EM square and stroke thickness can significantly affect legibility. In Japanese in particular, multiple writing systems coexist—kanji, hiragana, katakana, and Latin—so balancing their sizes and visual relationships is essential, along with considerations for vertical writing. Latin letters also require specific adjustments to harmonize with Japanese text; they cannot simply be designed in the same way as pure Latin typefaces.

Q: Latin typefaces often rely on open apertures (e.g., in c, e, s) and balanced spacing to enhance legibility. In more intricate or connected scripts, such as Arabic or Indic, what design strategies are essential to maintain internal clarity and prevent strokes or character components from merging visually?

A: For example, our Latin universal design typeface Clarimo UD PE has been expanded into multiple scripts. In Clarimo UD Arabic, we focused not only on distinguishing similar shapes but also on eliminating ambiguity in similar counter forms to avoid misreading. Maintaining open counters—similar to Latin design principles—also contributes to improved legibility. Reference: Collaboration with Rosetta (<https://rosettatype.com/services/Multi-script-Clarimo-UD-for-Morisawa>)

Q: Latin text is read letter-by-letter and word-by-word. For CJK typefaces, what is the fundamental visual unit a reader recognizes, and how does this affect the way Morisawa approaches the design of a typeface and legibility testing compared to Latin?

A: Kanji vary widely in complexity—ranging from simple characters like “一” to highly complex ones such as “鬱”. Because of this structural diversity, we place great importance on the balance between radicals and the distribution of internal space within each character.

Q: Variable Fonts allow easy manipulation of Weight, Width and other axis for fonts. What is one unique functional benefit or application of using Variable Fonts for CJK typefaces that is essential for multilingual or accessible digital environments?

A: CJK fonts contain a vast number of characters, and traditionally multiple weights had to be prepared individually. Variable Fonts allow continuous adjustments of weight, width, and contrast, enabling optimization for different user environments. They can also help reduce file size for web and embedded applications, giving users more flexibility. However, balancing aesthetic consistency with full functional interpolation can be challenging—ensuring no visual distortion across the design space is extremely demanding, and only certain design genres are currently practical for Variable Fonts. Reference: Variable Fonts (<https://www.morisawa.co.jp/culture/dictionary/15410>)

Q: With increasing globalization and multilingual communication, how does Morisawa anticipate the balance between global consistency and local specificity in type design, especially for accessibility and inclusive reading experiences?

A: Achieving the optimal balance is extremely challenging, but at Morisawa we collaborate with partners who possess expertise in local script development. While maintaining consistency across the type family, we strive to respect each script's regional typographic environment and cultural characteristics. Reference: Collaboration with Rosetta (<https://rosettatype.com/services/Multi-script-Lutes-UD-for-Morisawa>)

Q: How does your approach change from working on Latin to designing a multi-script project at Morisawa, what is the most important shift in your approach?

A: As mentioned in the answer to question 5.

Q: What are the key differences between evaluating legibility for Latin typefaces and for scripts such as Japanese, Arabic, or Devanagari?

A: We currently do not have publicly available information on this matter, so we kindly ask for your understanding.

Q: What are there particular metrics or user-testing methods that you use when designing CJK, Indic, or other complex scripts or minority languages? What are the biggest difficulties you encounter whilst doing user-testing?

A: We have conducted experiments and evaluations using methods such as Scheffé's paired comparison test. Reference: Comparative research report on the legibility of Morisawa multilingual fonts (UD Shin Go Hangul, UD Shin Go Simplified Chinese, UD Shin Go Traditional Chinese) (https://www.morisawa.co.jp/fonts/udfont/data/UDFontResearchReport_1804.pdf). We apologize that this document is not available in English; it would be helpful if you could translate it using AI tools.

Q: When designing typefaces for minority or less-digitally-supported scripts, what are the greatest challenges (technical, linguistic, or cultural) that you face during development and testing compared to more established scripts like Latin or Japanese?

A: One major challenge is the difficulty of gathering adequate resources on cultural elements and reading conventions for those languages.

Q: Considering your experience with type design across different writing systems, what factors most strongly influence legibility for readers with visual impairments? Do these factors differ significantly between alphabetic and logographic or syllabic scripts?

A: We currently do not have publicly available information on this matter, so we kindly ask for your understanding. Additional reference: "How Morisawa Fonts Are Made" (<https://www.youtube.com/watch?v=FJVT-ZLbp9Bg>)

Dave Crossland

Q: Google Fonts serves a global audience with varying degrees of visual health. In your view, is the type industry moving away from "intuition-led" accessibility (general rules like open apertures) toward "evidence-based" design that relies on clinical data and simulation models?

A: I don't think the type industry has been moving away from craft based design practice towards science, no. Mainly because reading science is validating existing knowledge established through craft, rather than really developing new insights; and because very few type designers work in direct collaboration with reading scientists when developing new type, and most of those cases are ex-post-facto studying designs after they are established. The impact was more about the explanations for why type designers do the things they do when they optimize a typeface for reading fluency.

Q: This research utilizes the VI-OCR pipeline to simulate the performance of low-vision readers. Given your work with font engineering and automation, do you see these tools that can accurately simulate the conditions of visually impaired users as a viable way to scale legibility testing for the thousands of fonts in the Google Fonts library?

A: Yes, I could see VI-OCR programs being integrated into QA software like github.com/googlefonts/fontquant and fonttools.github.io/fontspector. Somewhat related as an example, we have some [fonttools](https://github.com/fonttools/fontspector) (previously known as [fontbakery](https://github.com/fonttools/fontbakery)) checks already for the ISO 15008 automotive accessibility standard (<https://github.com/fonttools/fontspector/tree/main/profile-iso15008/src/checks/iso15008>).

Q: In the future we believe VI-OCR can help type designers and engineers find "Legibility Sweet Spots" across Variable Font axes (Weight, Width, Optical Size). How far are we from a web environment where fonts automatically "self-optimize" based on a user's clinical vision profile (e.g., an OS-level setting for Contrast Sensitivity)?

A: Well, there are a few different scopes in this question, so the answer is both 'zero distance, works today' and 'some years, perhaps never'...

For operating systems, they tend to move very slowly, and I am not yet aware of any OS with a setting for legibility factors directly – the closest settings are the ubiquitous system font-size settings. For web browsers (where the original early 90s term for those apps, 'user agents', is becoming relevant again in 2026 :) the 'reader mode' sometimes offers more settings, I think perhaps Edge has the best (<https://support.microsoft.com/en-us/topic/use-immersive-reader-in-microsoft-edge-78a7a17d-52e1-47ee-b0ac-eff8539015e1> + <https://support.apple.com/en-au/guide/safari/sfri32632/mac> + <https://support.google.com/chrome/answer/14218344?hl=en> etc). The web platform core technologies (html, css, js) were initially created with a vision of users being able to customize their page rendering (so called 'augmented browsing'), and browsers have offered extension APIs to enable extensions that assist non-developer users to do so for decades. Today I think <https://www.tampermonkey.net> and <https://stylebot.dev> are popular extensions like this, although I first came across this with <https://en.wikipedia.org/wiki/Greasemonkey> over 20 years ago. Augmentation of type settings can even be done with a bookmarklet (<https://davelab6.github.io/font-friend/>).

So, for the savvy web user, the kind of person who can give you, direct from the dome, a 5 minute speech on the differences between Chrome, Safari, Edge, Opera, Brave, Comet and Atlas – this is a reality today, if they be-

came interested in self-optimizing their ‘user agent’ based on their clinical vision profile... But for the everyday person, who walks their dog and mows their lawn, who buys a smart-phone and doesn’t mess with it too much, it seems far away to me. I suspect there will be a total revolution in how smartphones work before that.

Q: When commissioning or onboarding new typefaces for the Google Fonts library, what is the current protocol for evaluating legibility? Beyond technical checks, is there a formal metric that a design must clear regarding its legibility for the millions of users with visual impairments?

A: Generally, this is delegated to the type designer. The onboarding requirements do not get into this level of specificity – the first requirement is a broad one for ‘good quality’, which remains undefined further (<https://googlefonts.github.io/gf-guide/onboarding.html#new-fonts>).

There are two exceptions to this... First, that Google (via the Google Fonts team) was a member of The Readability Consortium (TRC, <https://thereadabilityconsortium.org>) and TRC evaluated the legibility of a number of Google Fonts in a variety of studies, and being able to recommend typefaces for inclusion in studies was one of the main value propositions for membership.

And second, the Google Fonts team was lucky to fill a User Experience Researcher (UXR) role with Hilary Palmén (<https://www.linkedin.com/in/hilarypalmen/>) in 2021, who has gone on to develop strong expertise in reading science. Hilary owns the evaluation of readability and legibility of type and typography for the Google Fonts team, within the larger parent group ‘Google Design Platform’ (<https://www.google.com/search?q=%22google+design+platform%22>).

Q: You’ve been a massive proponent of automated quality assurance through tools like FontBakery. Do you see “Legibility Simulation” becoming a modular, automated check in the font production pipeline, where a designer can’t, for example, commit code unless it meets a minimum visual accessibility score?

A: Yes, I do see “Legibility Simulation” becoming an automated check in font production pipelines, as part of the wider movement of popular computing careening towards artificial general intelligence – it is a simple case of Jevon’s paradox, where interest in, and the quality of, such simulations is rising exponentially, and the cost to obtain and run them is falling exponentially, so they will become created and used ubiquitously.

However, a restrictive integration test (where a project commit is rejected unless it passes above a minimum score) seems unlikely to me to be practical, since so many popular display typefaces must inherently score poorly on such checks.

In FontSpector (and previously Fontbakery), there are classes of check results, and I expect a low score will be a WARN, not a FAIL.

In the future I believe it would be interesting to integrate this technology directly into font editors (Glyphs/RoboFont/Fontlab). As a proponent of better design tools, how do you think real-time feedback on “Character Error Rates” would change the workflow of open-source type designers?

I agree such technology will be directly integrated into font editors too, as part of that ubiquitous trend I just mentioned, and you have an opportunity to lead the way here.

I foresee that such real-time feedback would most impact the workflow of novice type designers, irrespective of their posture towards the libre culture movement, since they don't have conscious knowledge of which type attributes affect readability and legibility.

I expect the InkTrap (typedesign.fba.up.pt) students are already too advanced for this, but that's where I'd suggest for you to start user-testing such a font editor plug-in :)

Q: With the rise of Vision-Language Models (like Gemini or Qwen2.5-VL), we can now ask AI to "describe" why it finds a letter difficult to read. Do you see a future where AI doesn't just score a font, but actually provides "Art Direction" to the designer on how to improve accessibility?

A: I'll be honest. I don't really see AI being used to assist type designers working with font editors, or even to generate fonts, very commonly. Of course it will be easy for someone acting as a type designer to do so. But rather, I see the entire edifice of typeface design, fonts, and everything else that was developed in the 70s for 'desktop publishing', becoming obsolete. DTP was largely making software systems to replicate, with more or less skeuomorphism, previous typographic/reprographic technology, going back past Gutenberg to the first Chinese printing presses. There in East Asia, people first developed the idea of typography, defined as "writing with pre-made letters". What hasn't changed since then and will not change is that people make documents, broadly defined. So when the marginal cost of generating letterforms falls exponentially, the need to pre-make letters falls away too, and the letters are made on-the-fly as the text is made and set all at once.

Today it is February 12, 2026, and tomorrow you may need to make a valentine's day party poster for the weekend. You can prompt the computer to make a poster, and it will ask you, or otherwise look up, or generate, the required information – the location, time, a hook headline, whatever, plus illustrations. If you don't like it, you tell it, and it generates however many more you want to see until you tell it you are done and ready to print, or send in whatsapp, whatever.

Inside that loop, the document's text is rendered, but there are no font files involved.

Gibson said "the future is already here but not evenly distributed," and <https://notebooklm.google.com> has been generating whole slide decks in that way, for many months now.

So yes and no, yes I do see a future where AI "provides art direction" to a type designer, and no, I don't see a future for type designers to provide direction to.

The 'general' word in 'artificial general intelligence' sounds a bit mundane, but a 'general purpose' intelligence system means one will solve any problem as well as an average person, if that average person could recall all existing knowledge.

As artificial intelligences advance inexorably towards general intelligence, they will "provide art direction" to the user for any design task with increasing efficacy. At least with typefaces, we don't have to consider how to neuter them.

Q: Human clinical trials are expensive, or time intensive and often limited to large corporations. Do you see open-source simulation tools like VI-OCR as a way to democratize high-level legibility research for independent designers, researchers and students?

A: Sure!

Q: If you could set one "Grand Challenge" for the next generation of type designers and engineers regarding visual impairment and digital reading, what would it be?

A: I'd suggest to run it like CASP, expecting the same ultimate result – complete automation, and a step change into refinement of the automaton. <https://en.wikipedia.org/wiki/CASP>

As we move toward using AI to simulate human impairment, there is a risk of "dehumanizing" the design process. How do we ensure that tools like VI-OCR remain a bridge to empathy for the visually impaired community, rather than just a way for designers to avoid talking to real people?

For actors who want to maintain a connection to the visually impaired community, tools like VI-OCR can be used alongside previous evaluation methods that involve talking to real people; and running multiple test strategies is essential for meta-analysis to nail in results/conclusions. However, since the cost of such methods doesn't fall as fast as more automated once, they will become rare in proportion.

Erik Van Blokland

Q: As a designer who builds his own tools, you've always looked for ways to make the "invisible" parts of type design visible. Related to this research, in the future I think it would be interesting to implement, in type design software (Glyphs/Robofont/Fontlab), a real-time feedback loop that shows "Legibility Error Rates" while a designer is drawing. Do you think having a "live score" would help a designer make better decisions, or would it be very limiting when designing typefaces?

A: It is difficult to judge the usefulness of an unwritten tool. But I'd encourage you to write a small prototype and see if it makes sense. To a very large extent the designer should be paying attention to such matters already. It helps to have a good proofing regimen, in which the shapes are exposed to all sorts of difficult contexts. Whether or not the type designer observes the right things, when looking at a proof, is an enduring question. But it gets better with experience. And it changes with each project. In general, a designer can not address any issues they can not see. A question might be "do you think such a score would help designers observe, or is it just making them guess."

Q: At TypeMedia, a huge part of the education is training the student's eye to recognize balance and rhythm. In the context of this dissertation, an AI (VI-OCR) toolset was trained to recognize letters under heavy blur. Do you think a machine can ever be taught to "see" the way a type designer does, or are we always going to be looking at two different things?

A: "Heavy blur" needs some.. qualification? Are you simulating with different Point Spread Functions, perhaps with some parameters connected to Zernicke polys? Ever since my eyes started to get bad I've paid attention to these transformations. And they are nothing like a Gaussian blur!

I start my class with a small practical experiment. At night, in a dark room, find a LED on a charger or something like that, a couple of meters away. Then try to observe the shape that dot makes on your retina and sketch it. Left and right will have different patterns, with and without glasses makes a difference. Each student will have a unique pattern. Comparing the patterns is really interesting because it shows that all people observe the world differently.

When the eye moves over the line, light reflecting off one letter enters in different parts of the eye at different times. When reading left to right, first in the right parafoveal view, then the center, then the left parafoveal view. So, a letter in word is transformed in different ways, several times. And the other eye transforms it in different ways again. With glasses, without glasses, outside, inside, barely awake or almost asleep. All that information hits the brain and some words are read.

This is how I analyse my own observations when I read my phone without my glasses. I'm sure everyone's experience is different and I don't want to draw general conclusions. But I'm hesitant to rely on a Gaussian blur as the central transformation when training your AI.

In that light I wonder how the OCR works. It is not a neutral, objective process. The code has been shaped by training data that can have all sorts of expectations built in. So, if your AI uses the OCR to find the moment where it stops to recognise letters - how do you know that moment is also significant for human observers?

I was wondering why you think the type designer is the person responsible to interact with this system? It might be more effective to add it to a design or typography tool. Later in the pipeline. Something that whispers to the designer that perhaps this text would benefit from a heavier weight and some tracking.

Q: In your work with FF Beowulf, you played with how much you could "randomize" or break a letter before it stopped working. If we use this proposed simulation software to find the exact moment a letter becomes unreadable, are we finding a useful limit for the craft, or is legibility too messy and subjective to be defined by a single "breaking point"?

A: The conclusion with Beowulf was there is no exact breaking point when a letter becomes unreadable. It is gradual. It is expressed by an increasing number of mistakes, a drop in comprehension, a drop in speed. At some point a reader gives up. But not everyone would give up at the same moment. But on the other hand, it is a cheap point to score because we already know that bad typography discourages reading. Mediocre typography will already be above this threshold and it still needs to get better! We want to practice good typography as a deliberate effort against that.

Q: You were a pioneer of interpolation with Superpolator. Most people use Variable Fonts for efficiency or web animations, but I'm looking at them as a way to possibly "tune" a font to a specific person's vision (like making it wider for someone with Glaucoma). Does this "adaptive" use of this technology interest you as a type designer?

A: I believe the reading experience improves when the reader has some agency over the process. Being able to move the device closer, or further, to accommodate. To hold the page in the light. Some control over the contrast, font size, font weight, line length, line spacing, all the things! are extremely helpful. Some digital reading experiences now offer something towards that.

The book as a physical object matters, a tight spine will hamper the long saccade back to the other end of the column. The weight of a book matters. Such design decisions are not necessarily difficult because it is really easy to see if it works. But a lot of designers do not take that time.

Why talk about typography? Because I think that most of the reading issues are *typographic* problems, not necessarily *type design* problems. Of course bad design decisions in a font can ruin the experience. But it is nothing compared to the number of ways careless typography can ruin the fun even when the fonts are fine.

Now if you have a digital reading app that is already set up with all the basic controls, then some sliders for finer interactions with the font might offer some relief.

Q: One of the tests I'm adapting is the "17ms glimpse" used by Larson with trigrams in the development of Sitka, where a reader only sees a flash of text. As a designer, when you look at a font for the first time, what is the very first thing your eye catches in that split second? Can we teach a tool to prioritize those same "survival features" of a letter?

A: As I tried to illustrate above, it is not a flash, the words ease up to the eye. The parafoveal saccade planner, the process that interprets the peripheral input and picks the spot to explore in the next saccade. How does that work? How is that visual input transformed, at the edge of our cornea / lens / retina? What information is available to the planner? That's what I would like to find out.

With that in mind, what does the flash represent? A single fixation? I can see a couple of problems if this is the only way you would judge the success of a letter. I understand it is relatively easy to measure, but I'm not sure how it relates to a specific part of a reading experience. So I don't know what the outcomes would mean.

If letters stand out, by excessive weight, or color, or movement, in the context of a paragraph, the parafoveal planner will go for the most interesting

nearby thing. It will pop down a line or two rather than down the line you were reading, if it stands out. So you will read the bold important word well before you read the rest of the preceding text.

Or, a different mechanism, but same result: try reading a paragraph next to an animated ad. Where do your eyes want to go?

A fresh look is not necessarily a first look. But sometimes it does help to put work away for some hours or even days and then see what stands out when looking at it again. But these would be most effective on paragraphs, proofs that show texture and quantity. Not just individual letters.

Q: You often emphasize the "black-and-white balance" and the rhythm of a typeface. For many visually impaired readers, letters "crowd" together and merge. If a tool suggests you should break your rhythm (like wider spacing) to help a reader, as a designer, would you follow the data or trust your sense of balance?

A: Ha! Well, spacing is perhaps the most intangible and trickiest attribute of type design. I can joke the only purpose of drawing letters is to have something to space. A bad design, well spaced will be more useful than a good design badly spaced.

I understand crowding, to some extent we all experience it if the text is too tightly spaced at small sizes. It is the typographer's responsibility to find the right font for the type size and line length, and apply tracking if necessary. It is the type designer's job to be aware of the range of sizes the type is aimed at. And proof it at those sizes and make sure it does what it claims.

Wider spacing is not the same as breaking the rhythm*. Crowding is not an issue of "too much rhythm" and I'm not sure how you would teach an AI to recognise the either state and their difference.

That said, wider spacing or tracking for smaller sizes is absolutely useful for all readers. Not all typographers and graphic designers bother, but the knowledge is quite easy to teach and understand.

Q: You created TypeCooker to give designers difficult constraints to solve. Could we treat "Accessibility" as just another TypeCooker recipe? For example: "Draw a lowercase 'e' that survives 50% blur." Does that sound like a fun design challenge, or is it too clinical?

A: Both "survives" and "blur" are so conveniently vague they do not mean anything. The "50%" offers a false suggestion of precision just because it is a number you can feed to a filter. Yes, the reading experience gets worse if a reader can't accommodate all the way. The reading might take a bit longer. It might propose a headache if it continues. But most of the time it is not a Gaussian blur!

Different letters will fail in different ways, when they get smaller. I would always want to judge a text, a collective, and not single letters.

Teaching design students about accessibility is important. For type it starts by being aware of the infinite number of different reading experiences. We draw stuff and hang it on the wall and observe the work from the other side of the hall. And that reveals useful things like "it is too thin" or "It is spaced way too tight" or "the counters are gone". This is not just for typography related to signage, but any experience in which the reader has no control over the geometry.

Q: The research shows that "perfect" or neutral shapes (like the very closed apertures in Helvetica) are the first to fail when someone can't see well. In your experience as a teacher and type designer, are we too obsessed with "clean" outlines at the expense of making characters truly distinct from one another?

A: There are no neutral shapes. There are shapes we have seen so often that our association with those shapes has become diffuse. By the same token there are also no perfect shapes. There are really good ones, and less good ones. And you will pick different ones from me. We can image all sorts of ranking systems to test one shape against another and make claims. But we also know that every new test would point to a new champion. The closed apertures of Helvetica are well known. There are many typefaces that do not have them. And many typeface specimens that point out they don't have them. Very well. They can still have crap spacing and terrible balance. Clearly it is not a single problem with a single solution. Making shapes "truly distinct" is only half of the job. Suppose all letters would have maximum difference from each other, nothing in common. There would be no baseline, there would be no x-height. Words and lines would refuse to form. There will be no reading.

Then you can argue "but of course there are some basic things that they need to have in common" and I would agree. Because it means it is not a binary state but a negotiation, and back to "balance". If letters share too much, information is gradually lost. If they share less to nothing at all, there will be no reading.

The point to this is that the overall appearance of the text likes to be even, with no extreme differences in weight, color, anything. Nothing to distract the planner. Then close up, in a fixation, the differences can manifest themselves. The challenge is to make shapes that have can evoke these different responses. There is a compelling argument for fitted text italics there as well.

Q: Usually, only big tech companies can afford to test how people read, with tests often taking a considerable amount of time. Do you think having these "validation tools" (VI-OCR) be open-source for all to use and improve, will help independent designers compete and create more legible and accessible typefaces, or will it just make everyone's fonts look the same? Are there any ethical considerations we must think about when implementing these tools in our own workflows?

A: Well, some of the readability research seems to be motivated for other reasons than science. It really depends on the tool and the kind of feedback it would offer. The risk of many AI tools is that they are opaque and fundamentally unaccountable. So if you make something that blurs a text and then runs OCR and counts the mistakes, you would learn more about the OCR algorithm than the letterforms. To have your work scored might invite the designer to guess what the machine then might like, rather than become critical and observe for themselves. It would add a gamification to the process that does not actually guarantee that the changes made to the design would improve reading for humans. Just that the score goes up. Typography is not difficult, it needs some patience and iterations.

That does not mean there is nothing to experiment with or work for. How about a tool that would visualise common visual impairments, applies some clever Zernicke math and image processing, half of the tool you propose really, and show that image to a designer. That could be helpful, but without the OCR and scoring. Because you don't know what the OCR is picking up on. Basically: ask a typographer over 40 to design a book / app because they will already observe the world in a totally different way than a 20 yr old with healthy eyes. Not in terms of style or ideas, just the stuff that hits the retina. Ask the 20 yr old to do the interaction and animations.

Q: You mentioned that you only have your "eyes and experience" to rely on. In a world full of AI and simulations, how do we make sure we don't lose the "human" side of the craft? Is there a risk that we stop trusting our own eyes because a computer gave us a "legibility score"?

A: The AI outputs I have seen and worked with are wrong in so many ways. I would not trust the score one way or the other. I'm weary of "but in the future this will get better" where there is no indication that it will.

Suppose a visually impaired person is given something to read that the computer says is legible, and yet the person still can't read it? How do you think that story will end? Who will be blamed? All the folks who had the skills, or could have acquired them, are gone. The knowledge of how the computer reached its conclusion was vague to begin with and has been forgotten and can't be reconstructed from the model itself. Someone will sign off on some work that will fail badly and someone else, down the hall, will get the blame.

What *would* be interesting to compute? What would be interesting to learn with computers that would improve reading? Precise measurements from orbscans and retinal scans, to build a mathematical 3d model, to do some ray tracing, to see what actually hits the retina? Well designed typography in books, devices, websites, which needs sensibly trained designers. More accessibility tools for adapting the reading experience for individual needs, and with a good understanding of what those needs might be. Or perhaps, with some eye tracking, if the device can observe the reader (there are definitely some creepy issues to deal with first) the text can respond in very subtle ways as a response to being read.

Mary Meürer

Q: In your research at UFSC, you approach design as an information facilitator. How do you see the role of typography in the autonomy of citizens with low vision in public spaces, and where does current design fail most in this regard?

A: Typography is fundamental to improve efficient information and ensure access for all people, independently of their reading condition. Obviously, it isn't easy to design for all, in particular to citizens with low vision, because their necessity of reading is variable and dependent on visual acuity. To develop accessible artefacts the designer needs to understand the context of the message, the situation of reading and the preferences or/and requirements of readers. All decision making must be based on these 3 factors, but this involves interesting research and conditions (time and investment) to make tests, evaluation and changes. Unfortunately, many times it is very difficult, or there is no commitment, to ensuring that the information is actually accessible. It should be noted that when it comes to publicly available information related to education, health, safety, and civil rights, access to information is essential.

Q: We know that conducting tests with human participants with visual impairments (VI) is complex, expensive, and sometimes exhausting for the participant. Do you believe that digital simulation tools, such as VI-OCR, can help democratize inclusive design for students and small studios who lack access to clinical labs?

A: Certainly, in this context, any resource that can support a more accurate use of typography is very relevant. Especially if the tool is built on a scientific basis, resulting in greater credibility. Just to give one example, I often recommend to my students the tools developed by Prof. Florian Adler's team, which help define font size and contrast. (<https://legibility.info/>) I recommend, of course, that user testing be prioritized whenever possible, but these tools are a valuable support.

Q: My research seeks to prove that this AI software toolset (VI-OCR) can simulate human reading difficulties. From an academic standpoint, what are the greatest risks of relying on computational simulations to dictate graphic accessibility standards?

A: An important issue is not to accept results without questioning them and without careful analysis. Biases exist and are very dangerous, and this applies to all uses of AI. Another point that needs to be noted is the general context, which involves content, reading conditions, and reader profiles, constituting a complex scenario. Adopting rules, whether they come from computer systems or even authors or researchers, without analyzing the entire context will always be a risk. An example I often give to students is a well-known table that indicates the recommended font size according to age. I alert them to the fact that it is not enough to know the size in points, as the x-height greatly affects the perception of size.

Q: In the protocol I develop, I reduced the number of trials to avoid fatigue among participants with VI. In your experience with user testing, how can we balance the need for statistical rigor with respect for the comfort and well-being of the low-vision participant?

A: Your concern is very relevant, as reading can be a very tiring and even a painful process for people with low vision. It is essential to ensure that they have a suitable environment, if possible one of their own choosing, and that

they can take the test at a pace they find satisfactory. I believe that the most important thing is that you are prepared to make changes to the procedures if necessary and that everything is properly recorded. Conducting a pilot test will also help in adjusting the protocol in advance.

Q: In the context of Variable Fonts being used to adjust type weight and contrast according to pathology (e.g., Glaucoma vs. Cataracts). Do you foresee a future where information design is not "static," but rather adaptable to the visual profile of each user via digital support?

A: Indeed, I even published an article on this topic with Professor Máira Wolozyń and student Dominique Auler. We tested how the axes of a variable font could contribute to accessibility. The intention was to verify whether it would be feasible initially. We have not yet delved deeper into this research, but if you would like to read the results of the article, it is available at: https://pdf.blucher.com.br/designproceedings/cidiconcic2023/92_651472.pdf In our article, we quote Garnham, P. (2023). (Variable fonts for a variable world. Available at <https://www.myfonts.com/a/font/content/variable-fonts-for-a-variable-world> Accessed on: May 16, 2023), which may be useful for your research.

Q: Often, legibility is treated as a purely optical factor. In instructional and educational contexts, how do you differentiate between "technical legibility" (the ability to distinguish letters) and "functional legibility" (actual comprehension of the message) for students with low vision?

A: It seems to me that this difference is equivalent to the concepts of legibility and readability. I think that both complement each other and that designers need to consider typography as a whole, type design, and design with types. I always remind my students that inclusive typography does not depend solely on choosing a font that prioritizes legibility. It is important to note this because we see some studies that focus on font characteristics and seem to consider that a font with certain formal aspects is what will ensure good reading, for example, for people with dyslexia. Of course, the font has a major impact, but it alone does not guarantee good reading conditions. It does not guarantee readability. Other factors are very important and involve layout, hierarchy, and the construction of the text itself, the choice of words, where, for example, the use of "plain language" comes in.

Q: As a professor, how do you approach teaching typography for accessibility? Is it possible to teach students to "see" like a person with low vision through simulators, or is direct experience with the user irreplaceable in a designer's training?

A: My priority is to make students aware of the importance of caring about the public. To really want to know more about their needs and how typography can help, or, if poorly applied, hinder the reading process and consequently access to information, which is everyone's right. To this end, I carry out activities where they need to read about the topic, look for tools that can help them, and try, really try, to talk to people who have low vision (or other disabilities, as I also work in other contexts). A simple exercise I do with them is to ask them to try reading while wearing glasses that simulate variations in the visual field according to certain pathologies. I point out, of course, that this is a simulation and does not replace testing with users, but it does a very good job of raising awareness. Putting yourself in someone else's shoes is not easy, but it is essential.

Q: Legibility can vary according to language and familiarity with words. When validating VI-OCR in Portuguese, do you believe there are particularities of our language (such as the use of diacritics) that represent a greater challenge for computational simulation than for the human eye?

A: Possibly. Diacritics are a great example, and I believe that the recurrence of ligatures can also have an impact.

Q: In Brazil and Portugal, we have certain standards for accessibility. Do you believe that the development of automatic validation tools could help update these standards for the digital environment more dynamically?

A: Absolutely. And in addition to facilitating updates, they could also facilitate customization. Although my studies have shown that there are many similarities between the accessibility demands of different audiences, we cannot ignore the particularities.

Q: By using AI to "simulate" a disability, we risk reducing the complex experience of visual impairment to a simple blur filter. How can we ensure that tools like VI-OCR are used to generate empathy and real solutions, rather than just "checking boxes" for technical project requirements?

A: I consider context to be fundamental. Associating the simulation of disability with a scenario that demonstrates the possible causes of this disability, how it impacts the reading process, and how it can compromise the experiences of people with low vision. Perhaps using persona building and storytelling techniques could be one way forward.

Fabian Harb

Q: Dinamo is known for pushing the boundaries of contemporary type, often embracing "uncanny" or highly characteristic letterforms. In a high-fashion or tech-branding context, is "accessibility" a constraint that limits creativity, or is it a technical challenge that Dinamo embraces as part of the "contemporary" look?

A: It totally can be both: Sometimes legibility, recognition, and accessibility are very conscious topics and requirements in (branding) projects – while other times a commission is much more about a certain expression, emotion and attention, and almost more close to creating an imagery than developing a tool? But type design in general is quite a special design task: On a zoomed-out level, there are tons of restrictions as an "a" always needs to look like an "a" to function. But more zoomed-in, there are a ton of possibilities as an "a" shape can be drawn in super many ways. And even in commissioned (custom) font projects, non-typepeople often lack a precise enough vision and vocabulary to dictate their exact wishes – so often we can bring in our own ideas, interests and preferences. And of course, by running a foundry we're aiming to create tools that are catering to a large range of tastes and tasks – hence creating letterforms with good legibility, recognition, and accessibility is an important part of it!

Q: Dinamo has a lot of custom tools, like those available inside Dinamo Darkroom. My research utilizes a VI-OCR pipeline, proposed by researchers at the Johns Hopkins University, to simulate low-vision failure. Do you believe that "automated validation" tools can have a place in the modern foundry's workflow, or is the "designer's eye" still the ultimate tool?

A: At Dinamo we're generally quite interested in technology and learning, using or developing new or proprietary tools and all – so, without knowing much about this VI-OCR pipeline, I am sure there are perspectives and projects in/for which this tool can be an interesting or exciting or helpful ingredient! What usually interests us a little less, is when tools become too closed, and automate (design) processes too extensively and do not really allow for other ingredients, project-specific inputs or even personal preference anymore?

Q: You've been pioneers in the Variable Font space. I am investigating the potential of using Variable axes to "tune" type for specific eye pathologies like Glaucoma or Cataracts. Do you see a future where "Accessibility Axes" could be a standard feature in some of your retail releases, allowing users to adjust the font to their own vision?

A: That could be super interesting and exciting of course! Not 100% sure if an axis is the right solution and whether these challenges and optimisations naturally translate to setups with axes and visual developments/movements/changes between masters and extremes, etc ... but the idea that a typeface can have features and functionalities "baked in" already, which in moments of need simply get activated or accessed or amplified, sounds awesome and super logical and promising to us of course!

Q: When you design a custom font for a global brand, do you test for "split-second" recognition or any other accessibility metrics, or is the focus purely on the visual identity and rhythm?

A: It totally depends on the client and commission of course: For a project like designing a new font for Spotify that needs to be used across many products and platforms and will be read by millions of people, of course accessibility is a huge concern and testing and optimising shapes an integral

part of the design development. In smaller frameworks though, often it is more difficult to allocate budget and time to care about it the same way. It usually comes down to the accumulated experience and knowledge and the dedication and commitment of the involved parties.

Q: I want to understand exactly which characters (like 'a' vs 'o') collapse into one another under simulated visual impairment conditions. Would you find it useful to have a "stress-test report" for a new design that highlights these weak points before the font is published?

A: As said before: Generally, we totally feel that tools to test and analyse typeface designs on their legibility, or recognition, or accessibility, etc. are of great help and value of course! But we feel that with bigger availability and easier access to such tools, it also becomes more important to carefully evaluate what is the right context and moment to make use of it, and when it is actually not.

Q: Clinical legibility studies are usually reserved for giants like Microsoft. If my research proves that an open-source tool can accurately mimic a low-vision reader, do you think this empowers independent foundries like Dinamo to compete with the "scientific" marketing of big tech companies? How do you think these types of tools can be integrated in your creative process?

A: We can totally imagine that this knowledge and expertise will be more democratically accessible in the future and therefore empower smaller design studios and foundries to take on projects that before were monopolised by "giants" like Monotype, or Adobe, or Google or so – which can be great of course! But having done a couple of these bigger projects in the last years, we also came to learn how they usually come with huge communication, mediation and negotiation efforts – and that it is often quite hard to develop interesting design work in such contexts ... so, in the end it probably stays a bit depending on the individual interests and ambition of each studio and foundry, whether and how much such integration makes sense.

Q: Google Fonts and other platforms are pushing for higher accessibility standards. As an influential independent foundry, do you feel there is a responsibility to educate the next generation of designers that "cool" type must also be "readable" type? What lessons do you believe we still have to learn in that aspect?

A: In our opinion, it's not the "readable typefaces" that have an "uncool connotation" – but much more the actual places where they come from? We acknowledge of course that Google Fonts have started to offer quite well-produced tools for a lot of language contexts that historically have been underserved. But we also witness how their initiative is deflating the value of typefaces and undermining the development of interesting original new typefaces, as they frequently commission designs and designers that borrow aesthetically from already existing sources ... So, if we continuously want to see new exciting and well produced typefaces getting released, we need to direct support and money towards places where this is actually done. And even if an interesting and original typeface is showing weaknesses regarding its legibility, or recognition, or accessibility, that almost always is something that type designers and foundries are eager to learn about and happy to improve for their audiences.

Q: In many companies fairly "neutral" sans-serifs that often fail under low-vision conditions are still the "standard". Why do you think the design world is so attached to these "invisible" but exclusionary defaults, and how can Dinamo's more "vocal" typefaces help break that cycle?

A: Again this tendency less feels like a reflection of what company values and believes, but mostly like a confirmation that large parts of the brand world is afraid to differ and to take risks? If you'd ask them directly, I am sure that almost all of them would agree that their identities and communication designs should be as legible, and recognisable, and accessible as possible. Because that of course enlarges their potential audience and can boost their sales. But then when it comes down to actually following through on the logical consequences, their designs probably start to differ a bit from their mood-boards and north-stars and the competitors left and right – and that seems where these attempts often stop again already.

Q: One of my dissertation's proposals is to find a way to integrate VI-OCR in our creative process and methodologies as type designers. Would you use a plugin, or a feature that would provide "Low Vision Preview" inside editors like Glyphs? Would you value a real-time "Character Error Rate" score that changes as you move a node, or would that be too much "noise" in the creative process?

A: We'd love that of course! With certainty there is things that we could learn and improve as well still – but especially for conversations with clients we feel that such tools could be a very practical and valuable as well: To make them more comfortable and reassured that these things aren't scary – and that they can "push their envelope" visually AND tackle legible, and recognisable, and accessible at the same time.

Q: As we move toward a world of AI-driven interfaces and VR/AR, the "environment" of the pixel is changing. How do we ensure that the human experience of reading, especially for those with visual impairments, remains at the center of the technology we build?

A: A little while back we were – half jokingly – wondering if maybe the font-files of the future will also contain voices, and not allow designers and readers to define the visual connotation of written contents; but also the "style of sound" with which their messages come through acoustically. In some media outlets you can already find automated voices offering an alternative to reading their contents – hopefully such tools become even more mainstream in the future and allow for equal access to information, no matter if you read or not. If AI/VR/AR can amplify or enhance these experiences, we're all for it. But it can start way more direct and small we feel. Let's better not put this topic and authority into the hands of big tech only!

Ricardo Santos

Q: As a prominent figure in Portuguese typography and a professor at ESAD, cr, how do you see the current state of "Inclusive Design" in Portugal? Are we moving toward a more evidence-based approach, or is accessibility still treated as a secondary technical requirement?

A: No sentido geral o Design inclusivo ainda é um aspeto secundário em Portugal, obviamente existem pequenas evoluções tanto em projetos profissionais mas também ao nível académico através de dissertações de mestrado e doutoramento.

A (translated): In general terms, inclusive design is still a secondary aspect in Portugal. Obviously, there have been small developments both in professional projects and at the academic level through master's and doctoral dissertations.

Q: In your own type design practice, you often work with very specific weights and structures. When designing for clarity, what are the "non-negotiable" anatomical features you focus on to ensure a typeface doesn't collapse under poor reading conditions?

A: Existem alguns fatores determinantes na consistência de um tipo de letra, primeiro é importante perceber qual é o meio ou output do qual irá ser utilizado (impressão ou digital). Relativamente aos aspetos técnicos que podem interferir com performance num determinado tipo de letra: 1) Normalmente as proporções horizontais (largura) podem variar, num contexto mais abrangente de leitores, os casos mais extremos (condensados ou expandidos) podem criar mais problemas na velocidade de leitura. 2) Altura de (x) tem um forte impacto na leitura. 3) O peso e contraste tipográfico é talvez dos aspetos mais importantes, porque é através destes dois parâmetros, que nivelamos a cor tipográfica do tipo de letra conforme o peso óptico desejado. 4) A entrelinha desenvolvida e aplicada num tipo de letra deverá estar pensada para o tamanho seu óptico (legendas, texto, sub-títulos, display). 5) Relativamente ao desenho e anatomia é preferível não haver formas exageradas ou acentos muito grandes. 6) O espaçamento dos caracteres também pode influenciar a leitura, não só para percepção de letra a letra como no reconhecimento da palavra. 7) Quando trabalhamos para leitores com problemas de leitura ou percepção, temos que perceber que alguns dos cânones tipográficos clássicos, podem não funcionar e temos que estar abertos ao desconhecido.

A (translated): There are a few determining factors in the consistency of a font. First, it is important to understand the medium or output for which it will be used (print or digital). Regarding the technical aspects that can interfere with performance in a given typeface: 1) Normally, horizontal proportions (width) can vary. In a broader context of readers, the most extreme cases (condensed or expanded) can create more problems in reading speed. 2) The height of (x) has a strong impact on reading. 3) Typographic weight and contrast are perhaps the most important aspects, because it is through these two parameters that we level the typographic color of the font according to the desired optical weight. 4) The line spacing developed and applied to a font should be designed for its optical size (captions, text, subtitles, display). 5) In terms of design and anatomy, it is preferable to avoid exaggerated shapes or very large accents. 6) Character spacing can also influence reading, not only for letter-by-letter perception but also for word recognition. 7) When working for readers with reading or perception problems, we have to realize that some of the classic typographic canons may not work, and we have to be open to the unknown.

Q: The Portuguese language relies heavily on diacritics (~, ^, ', ` , ç). I've noticed that these are often the first elements to "wash out" under simulated blur. How do you approach the design of diacritics to ensure they remain functional for readers with low visual acuity without becoming a distraction?

A: Não tenho muitas informações sobre o contexto dos acentos (nomeadamente na língua portuguesa). Os acentos num contexto normal, para corpos de letra mais pequenos (legendas, texto) são relativamente maiores e mais fortes, no caso de tamanhos grandes (display) a opção é reduzir a dimensão, para terem menos presença junto das letras. Possivelmente a sugestão será para aumentar a dimensão dos acentos onde o seu desenho deverá ser bem diferenciado e reconhecível. Também existia a possibilidade de unir os acentos com as letras, mas essa opção não recomendo, porque normalmente produz resultados mais estranhos.

A (translated): I don't have much information about the context of accents (particularly in Portuguese). Accents in a normal context, for smaller font sizes (captions, text), are relatively larger and stronger; in the case of large sizes (display), the option is to reduce the size so that they are less prominent next to the letters. Possibly the suggestion would be to increase the size of the accents where their design should be clearly differentiated and recognizable. There was also the possibility of joining the accents with the letters, but I do not recommend this option, because it usually produces stranger results.

Q: At ESAD.cr, you mentor future designers. How do you teach students to balance the "artistic" side of type design with the "service" of legibility? Do you think tools like a "Low Vision Preview" integrated into the design software would change how students learn to draw?

A: A grande maioria dos alunos (licenciatura) nas aulas de design de tipos de letra, centram-se mais nos aspetos estéticos ou decorativos do que no funcionalismo. Normalmente os alunos de mestrado estão mais aptos para projetos com resultados mais interessantes devido à investigação. Na realidade o design é baseado em aspetos simples, por isso quanto mais elementos diferentes incorporas num tipo de letra, mais difícil é a sua homogeneização e harmonização formal. Penso que é importante sensibilizar os alunos sobre o tema "Low Vision Preview", o facto de futuramente poder haver ferramentas de teste incorporadas nos programas gráficos, seria uma excelente forma de promover o tema.

A (translated): The vast majority of students (undergraduate) in typeface design classes focus more on aesthetic or decorative aspects than on functionality. Master's students are usually better suited to projects with more interesting results due to their research. In reality, design is based on simple aspects, so the more different elements you incorporate into a typeface, the more difficult it is to achieve homogeneity and formal harmony. I think it is important to raise students' awareness of the topic of "Low Vision Preview." The fact that there may be testing tools incorporated into graphics programs in the future would be an interesting way to promote the topic.

Q: My research proposes that VI-OCR can potentially simulate visually impaired people in legibility testing. As a designer who values the "craft," do you feel comfortable trusting a machine's "Legibility Score," or should the final validation always come from a human eye? Do you have any ethical worries of a tool like this being integrated into your workflow?

A: Estas ferramentas podem desempenhar um papel importante na produção de dados, mas numa fase inicial é sempre bom haver uma supervisão humana para a validação dos mesmos. Estas ferramentas para funcionarem

de forma mais rigorosa e detalhada, terá que ser ensinada e atualizada conforme casos padrão existentes.

A (translated): These tools can play an important role in data production, but in the early stages it is always good to have human supervision to validate the data. For these tools to work more accurately and in greater detail, they will need to be taught and updated according to existing standard cases.

Q: You've seen the transition from static to Variable Fonts. I am investigating how to use Variable axes to "tune" type for specific eye pathologies (Cataracts, Glaucoma). Do you believe this level of personalization is the future of typography, or should we keep striving for a "Universal" accessible font?

A: No futuro e até na atualidade é possível ter uma melhor personalização dos diferentes sets, hoje em dia através das open type features é possível alterar os diferentes conjuntos de letras pretendidos (ex: stylistic characters). Neste momento existe uma condicionante a ter em conta, os tipos de letra dos quais utilizam as open type features tem um funcionamento reduzido, só funcionam em programas gráficos (ferramentas de design), muitos dos programas gerais não comportam a utilização desta tecnologia, para isso o mais recomendado é separar o tipo de letra em diferentes fontes tipográficas individualmente com cada conjunto. O mesmo se aplica à tecnologia das variable fonts que também esta limitação técnica, funciona em poucos programas e browsers. Se este tipo de letra em formato VA, for utilizado num ambiente controlado (computadores, programas, equipados com os as ferramentas adequadas esta problema talvez não se coloque. Muitas vezes o termo Universal é utilizado, mas ainda hoje existem estas pequenas condicionantes tecnológicas. O mesmo se aplica se o tipo de letra por exemplo for disponibilizado ao público (users target) em grande escala, muitos dos utilizadores possuem ferramentas mais arcaicas.

A (translated): In the future, and even today, it will be possible to better customize different sets. Nowadays, through open type features, it is possible to change the different sets of letters desired (e.g., stylistic characters). At this time, there is one constraint to consider: fonts that use open type features have limited functionality, only working in graphics programs (design tools). Many general programs do not support the use of this technology, so it is best to separate the font into different typefaces individually for each set. The same applies to variable font technology, which also has this technical limitation and works in only a few programs and browsers. If this type of font in VA format is used in a controlled environment (computers, programs equipped with the appropriate tools), this problem may not arise. The term "universal" is often used, but even today these minor technological constraints still exist. The same applies if, for example, the font is made available to the public (target users) on a large scale, as many users have more archaic tools.

Q: When you are spacing a font, do you think about how the letters interact at the limits of perception, or is the "rhythm" of the font something that can only be judged under perfect conditions?

A: Para os tipos de letra para tamanhos mais pequenos (legendas ou texto) o espaçamento pode ser ligeiramente mais aberto, porque em pequeno existem uma maior necessidade das letras poderem respirar visualmente. Nos tamanhos display o espaçamento tende a ser mais apertado, evitando sempre a colisão dos traços ou das serifas. Muitas vezes sei que alguns dos meus alunos têm dificuldade em perceber o que é o espaçamento original do tipo de letra, com o espaçamento que é aplicado numa caixa de texto (conforme o tipo de alinhamento poderá influenciar o espaçamento do texto

original).

A (translated): For smaller font sizes (captions or text), spacing can be slightly wider, because smaller fonts need more room to breathe visually. In display sizes, spacing tends to be tighter, always avoiding collisions between strokes or serifs. I often find that some of my students have difficulty understanding what the original spacing of the font is, with the spacing that is applied in a text box (depending on the type of alignment, this can influence the spacing of the original text).

Q: Many commercial clients still demand "neutral" or "clean" aesthetics (like neo-grotesques) that research shows can be exclusionary for the visually impaired. As a consultant and foundry owner, how do you convince a client to choose a more "idiosyncratic" but legible design over a "trendy" but illegible one?

A: Neste momento não realizo trabalhos de design gráfico do qual tenha este problema de convencer os clientes a usar determinado tipo de letra. Muitas vezes os clientes (Portugal) querem usar o que todos usam, ou a manter o uso da Helvética como um padrão corporativo moderno. Muitas vezes as opções mais neutras tem tendência a prologarem mais no tempo e envolverem alguma intemporalidade gráfica, as opções mais arrojadas ou de modas acabam por ter um tempo de vida mais resumido (a história do design gráfico no séc XIX e XX está cheia de modas, mas são poucas as que ainda permanecem no ativo (o modernismo é um deles).

A (translated): At the moment, I don't do any graphic design work where I have this problem of convincing clients to use a certain typeface. Often, clients (in Portugal) want to use what everyone else is using, or to stick with Helvetica as a modern corporate standard. Often, the more neutral options tend to last longer and involve some graphic timelessness, while the bolder or more fashionable options end up having a shorter lifespan (the history of graphic design in the 19th and 20th centuries is full of fads, but few of them are still in use today (modernism is one of them)).

Q: One goal of my dissertation is to validate an open-source toolset (VI-OCR) for independent designers. Do you think a more "scientific" approach to testing would help the Portuguese type design community gain more visibility and competitive edge internationally?

A: Sim penso que é muito benéfico a utilização de processos, metodologias de um modo mais científico de forma a parametrizar todo o trabalho desenvolvido, só assim é que percebemos a causa e efeito, na decisão de determinadas opções num projeto. Eu diria que o espectro de difusão do projeto deverá ser internacional, não é pela questão de visibilidade do projeto, mas pela interligação global do tema e partilha de conhecimento de forma centralizada.

A (translated): Yes, I think it is very beneficial to use processes and methodologies in a more scientific way in order to parameterize all the work developed. Only then can we understand the cause and effect in the decision-making process for certain options in a project. I would say that the project's dissemination spectrum should be international, not for the sake of visibility, but because of the global interconnectedness of the topic and the centralized sharing of knowledge.

Q: You've seen type evolve across different mediums. Does the responsibility for accessibility lie with the person drawing the curves (the designer), or is it the responsibility of the software and OS developers to fix the "broken" vision of the user?

A: A responsabilidade recai principalmente nos programas e criadores, nomeadamente na criação de uma maior acessibilidade tecnológica. O designer de tipos de letra, muitas vezes não consegue resolver os problemas engenharia na incompatibilidade dos programas, deverá criar os caracteres, conforme as diferentes descrições dos problemas visuais e suas necessidades básicas. Para o desenvolvimento um projeto desta natureza, são necessárias várias valências e meios tecnológicos integrados. Este tipo de projeto de um nível de complexidade alto e muitas vezes os designers de tipos de letra, não têm tempo e capacidade para investigar isoladamente um meio tão complexo e multi-disciplinar.

A (translated): The responsibility lies mainly with the programs and creators, particularly in terms of creating greater technological accessibility. Font designers, who are often unable to solve engineering problems related to program incompatibility, must create characters according to different descriptions of visual problems and their basic needs. Developing a project of this nature requires a range of skills and integrated technological resources. This type of project is highly complex, and type designers often lack the time and capacity to investigate such a complex and multidisciplinary field on their own.

Maíra Woloszyn

Q: Your work often focuses on how design facilitates the use of information. From your perspective, is the type design industry currently doing enough to address accessibility, or are we still focusing too much on the aesthetics of the letterform rather than its performance for low-vision users?

A: There are still no definitive studies that fully guide the development of typography for low-vision contexts. There are recommendations of varying natures, which are sometimes scattered and fragmented, potentially limiting knowledge. Regarding the practices of the type design industry, I perceive a concern among type designers to develop typefaces that are as legible as possible. However, I don't see this as necessarily the main focus of typographic projects. I believe that most focus primarily on the aesthetics of the letters, but always striving for legibility. I understand that there is a concern to make reading easier for the largest number of people possible. Even so, it is worth considering that generalizing the audience is often complex, since legibility issues for low-vision audiences must take into account not only formal aspects of the typefaces, but also the user's profile, the type of information, and its context. There is a difference between dealing with a frequent reader of long, narrative texts and an infrequent reader who seeks to identify information in signage or short informational materials.

Q: My research seeks to validate VI-OCR as an accurate way to simulate people with visual impairments for legibility testing. Given your background in research, what do you believe are the essential "human factors" that a computational simulation must capture to be considered a valid tool for inclusive design?

A: I believe that a tool that provides the perspective of a user with low vision can contribute to understanding the studied context, such as color simulators in design. However, it is important to emphasize that simply obtaining the same visualization as a person with low vision excludes the perception of this audience, their context, and possible difficulties that may exist in their interaction with the text/typography.

Q: I am using Kevin Larson's 17ms trigram protocol to measure legibility thresholds. As someone who studies how people process information, do you think testing letters in isolation (trigrams) is sufficient, or does the "context" of a word provide a level of accessibility that a machine simulation might miss?

A: Identifying letters in isolation can be a good starting point. However, I believe that context is very relevant and should be considered in the process of providing guidelines for inclusive typographic design.

Q: My investigation shows that a font optimized for "Standard" vision can fail significantly for someone with Cataracts or Glaucoma. Do you think the future of Information Design lies in "Universal Design" (one font for all) or in "Pathology-Specific" typography that adapts to the user's clinical profile?

A: I understand that a universal typeface is a complex and difficult idea to achieve, as each case of low vision has its own particular contexts, along with other issues that can hinder reading. Furthermore, typefaces that aim for greater "cleanliness" often sacrifice space for details that bring personality to the typeface. However, variable font technology can contribute to adjustments in a typeface that can expand the possibilities within the same typographic design, as explored in this article: https://pdf.blucher.com.br/designproceedings/cidiconcic2023/92_651472.pdf

Q: Variable Fonts allow us to manipulate weight, width, and optical size dynamically. How do you see the role of these "fluid" systems in creating more inclusive digital environments, particularly for users who need real-time adjustments to contrast and density?

A: I believe that variable fonts are a practical resource for improving the accessibility of typefaces according to the user's needs in their specific context, due to the fine adjustments provided by the technology.

Q: In terms of User Experience, how critical is "Character Disambiguation" compared to other factors like line length, leading, and color contrast for a low-vision reader?

A: Identifying characters is important for comprehension. However, it is worth noting that understanding words and phrases often carries more weight depending on the context (a book, a quick informational piece). Therefore, I believe both situations should have equal weight in type design.

Q: Some critics can argue that using AI to simulate a disability "dehumanizes" the user. However, I argue it can be an "Empathy Tool" for designers, by making accurate simulations integrated in our design process, not excluding human testing as a final stage, just changing what we test, when we test, in order to reduce testing costs. Do you think seeing a real-time "Legibility Score" or a "Blur Simulation" in a design tool can help designers build more empathy for the visually impaired community?

A: Yes, I believe these tools provide the type designer with a perspective closer to the user's reality. This helps to understand their difficulties and limitations, which can lead to a project that has more potential to meet the audience's demands.

Q: Clinical trials are expensive and often inaccessible to independent designers in Latin America, Portugal and other regions of the world. Do you believe that open-source tools like VI-OCR can democratize the ability to produce high-performance, inclusive typography outside of major tech hubs?

A: Yes, open and easily accessible tools are necessary and can contribute in an easy way to the development of highly complex projects.

Q: We see many public health and transit systems using "neutral" neo-grotesque fonts. My research suggests these are often the least accessible. Why do you think Information Design remains so attached to "neutrality" when the evidence points toward more "idiosyncratic" shapes being safer for reading?

A: I understand that the adoption of "neutral" fonts stems from the idea of universality and simplicity that began in modernism. Furthermore, I believe that this choice, in these contexts, can also be associated with the need for quick recognition of information, and therefore, opting for typefaces with few ornaments.

Q: As someone who works at the intersection of practice and academia, how do we bridge the gap between "Lab Results" (like OCR data) and the "Real World" (the person reading a medicine bottle or a transit sign)? How do we ensure this data actually changes how things are designed?

A: I believe this convergence happens when "worlds" truly meet. Listening to real people, conducting readability and comprehension tests with the public, and seeking to understand (also through tools) the realities that will be worked on, is the way we have to unite "lab results" with the "real world".

Furthermore, seeking to disseminate scientific data and make it accessible (both in terms of access to it, but also in simple language) to those who develop projects aimed at these consulted and understood audiences allows for the development of increasingly appropriate information systems.

Octavio Pardo

Q: Your work often deals with optical sizes and the fine-tuning of character shapes for specific environments. In your experience, which geometric features (apertures, stroke contrast, x-height) are the first to "collapse" when a reader has low visual acuity, and how does that influence your initial sketches?

A: There are probably several valid answers to this question depending on the specific visual impairment we're talking about, and above all on context, since almost any typographic parameter is size-dependent — that is, it must be adjusted according to the visual angle at which it is perceived. The only parameters we know consistently benefit reading performance are low contrast and generous apertures. The features that tend to collapse first are precisely insufficiently open apertures and the internal counters of dense characters such as "a", "e", and "g". In practical terms, this knowledge influences my process by forcing me to ensure that my initial sketches are deliberately wider and more spaced than what would feel natural inside Glyphs or any font editor. That usually provides a solid starting point.

Q: VI-OCR was developed to simulate several types of visual impairments. As a type designer, do you currently use any digital filters or methods to test your outlines, or do you find that these simulations are often too detached from the reality of human reading?

A: I don't use any filters beyond the blur effect in Glyphs, which is quite practical. However, we have very limited information about how to calibrate it, which restricts its rigorous application. More broadly, most digital simulations tend to oversimplify perception. They can approximate degradation, but they rarely capture the cognitive component of reading.

Q: As a consultant for Google, you see how fonts are consumed by billions. When onboarding a new typeface for such a massive platform, what are the primary legibility "red flags" you look for, and do you think we are close to having an automated standard for accessibility at that scale?

A: We are nowhere near having an accessibility standard at a typographic level because we lack meaningful structural descriptors for how typefaces should be adjusted. We use terms like "width," "bold," or "slanted," which make conceptual sense, but in terms of letter architecture they mean absolutely nothing. And interestingly, the discipline does not seem particularly interested in generating or adopting more precise descriptors. As for red flags, there are fewer than one might imagine. Screens have become so refined that they simplify much of our work. Still, one must always consider monitor proportions, contrast, apertures, counters, weight distribution, and spacing.

Q: Your PhD involved deep research into the legibility of typefaces under specific constraints. How has your academic background influenced your professional consulting work, particularly when you have to justify "unconventional" design choices for the sake of better reading performance?

A: I'm going to ruin the mystique of having a PhD in legibility. Precisely because of the lack of a rigorous scientific typographic taxonomy, there is no strong body of literature capable of supporting particularly risky design decisions. Since there are not many highly skilled type designers conducting experiments, many academic papers rely on typefaces whose design quality and performance should, at the very least, be considered questionable. Ultimately, what contemporary legibility science has contributed most to type design is methodology: a solid framework to question and test ideas. But the generalization of those results is another story entirely.

Q: You have worked extensively with modern font technology. Do you see Variable technology as an opportunity to make typefaces more accessible? Do you think in the future of type design we will see typefaces that can be customized by the user become more common and a “standard”?

A: I’m not sure whether we will see this applied in a truly practical and dynamic way, but I am convinced it is the best opportunity we have. What is almost tragic is that we already have the tools: devices with cameras sensitive enough to track eye movement, highly flexible variable fonts, and enough computational power to generate instances on the fly and test their performance. We also have the knowledge to do it. The bottleneck is not technological, it is economic.

Q: In psychophysics, the staircase method has been documented as a way for finding the threshold of perception. As a researcher, do you believe that a computational model (like an OCR engine) can truly replicate the “learning curve” of a human participant in these trials, or is the machine’s logic too linear to simulate human cognitive fatigue and adaptation?

A: It will depend on how the model is trained. The more behavioral data from human participants we collect, the more accurately a system could replicate the human performance curve (improving at first, stabilizing, and eventually declining due to fatigue). The theoretical possibility is there. The real difficulty lies in the programmer’s ability to gather sufficiently rich and representative data to train such a model.

Q: With models like Gemini 1.5 Pro, we can ask an AI to describe exactly which part of a glyph is causing a “recognition error.” Would you find it useful to have a tool that gives you a “Legibility Score” or a “Heatmap of Confusion” during the design process, or would it restrict your creative intuition?

A: If the tool comes with a clear and convincing rationale explaining how its recognition or scoring engine has been constructed, I believe it could be an excellent resource. Without methodological transparency, however, it risks becoming a decorative metric rather than a meaningful one.

Q: We often see “neutral” fonts like Helvetica being used in public information, yet they are notoriously difficult for low-vision readers. Why do you think the industry is so hesitant to adopt more “distinctive” or “idiosyncratic” shapes (like those in Luciole or Sitka) even when the data shows they perform better?

A: Ha! For the same reason that neutral sans serifs are the best sellers of every foundry. Bold decisions grounded in strong criteria are rare. The knowledge base that “the majority” relies on to make design decisions is often extremely limited and shaped by external influence rather than internal judgment.

Q: Clinical trials with visually impaired participants are expensive, time-consuming and rare. If my research proves that an open-source AI script can accurately simulate these human trials, do you think it will lead to a more inclusive generation of independent type designers? How can we integrate this type of toolset in the type design process?

A: Absolutely. If the tool is robust, I am confident its integration would be natural and almost spontaneous, after some degree of evangelization. That said, the ethical implications are relevant. What is the value proposition of such a tool? If the goal is to empower designers without access to expensive trials, it is extraordinary. But it should not aspire to replace experiments conducted with human participants. If our decision-making ecosystem becomes entirely dominated by AI systems, we move one step closer to being

replaced by them in those very decisions. It is crucial to remember who we design for and not lose focus on the user.

Q: As someone who has collaborated with major foundries and tech giants, where do you think the responsibility for accessibility lies? Should it be the designer's duty to ensure a font works for everyone, or should the software/browser be responsible for adapting the type to the user's needs?

A: It depends somewhat on the type of project. In general, I would say both parties share responsibility, but ultimately the responsibility lies with the final decision-makers: the clients. Think about it this way. Imagine you are redesigning a brand font for a large software company. At some point, a design decision must be made regarding a particular feature: one direction moves away from the brand tone but improves accessibility; the other resonates more strongly with the brand but results, for example, in more closed counters. I can assure you that if you ask the designer, they will almost always choose the first option. But the client? In most cases, they will choose the second.

Gerry Leonidas

Q: You have spoken extensively about "minimum quality" in typeface design. In the context of visual impairment, where the margin for error is almost zero, should "accessibility" be treated as a separate category of design, or is it simply the logical extension of professional "minimum quality"?

A: If I recall when I was discussing "minimum quality" the context was in relation to genre (which is set- member based description, as opposed to a categorisation) and subjective experiences of reading. I am pretty bullish that reading cannot be separated from the social dimension of both the act of reading, the lived experience of language (memory and more broadly textual experience), and the social context in which reading takes place. What works in one environment, does not for the same reader in another. With that said, accessibility has to be seen as a broader condition, including not only typographic parameters, but linguistic, and contextual (from situational to emotional). All these factors compile a multi- dimensional space in which reading happens, and through which every reading experience is navigated. Absolutely for specific visual conditions you can investigate typeface and typographic factors. But they are the most straightforward set of parameters, so should be seen as a starting point, not a final aim. So, to answer your specific question: yes, there is a multidimensional continuum within which different intensities of access to texts happen. To attempt to simplify these is only justified in order to carry out an experiment (in terms of time, resources, funding) and insofar the researchers are aware of the massive assumptions required. (This is reminiscent of the problems economic theories faced with "rational actors", if you want to go further.)

Q: This research uses a TrOCR (Transformer-based OCR) tool, created by researchers at the Johns Hopkins University, to predict human reading failures. Having seen decades of legibility research, are you skeptical of using a machine as a simulation for the human eye, or do you believe that a computer model can accurately reflect the structural limits of a letterform?

A: "Decades of legibility research" = very little that has stood the test of time. (See Ole Lund's PhD for a start in this discussion.) Huge problems with experimental assumptions, sample and participant biases, and very good candidates for entries in the Replication Crisis catalog – but saved because legibility research is relatively small and not mainstream enough. Not to mention that it is overwhelmingly Anglophone and North-Western, and any researcher who claims this doesn't matter is willfully blind. (BTW, one of our PhDs is the very first project looking at readability of Bengali in situ – can any western researcher in good conscience claim that any finding from Latin-based research is meaningful there?) Additionally, and speaking with the experience of discussing things with colleagues in our University's neuroimaging lab, everything was rewritten after MRI testing became widely accessible around 2009 or so – huge amount of research rewritten (and feeding into my point about the dimensions of "the act of reading", above). So, yes, sure. Simulate an eye all you want. What happens downstream, when whatever signals hit the brain? Or do you care only about the very narrow mechanics of shape distinction? (Because arguably shape identification is irrelevant in individual letters or shapes, if you want to inform comprehension.) So, what's the purpose?

Q: I am using Kevin Larson's 17ms trigram protocol to strip away linguistic context and test "pure" visual recognition. Do you believe that testing characters in isolation is a valid way to improve typeface design, or does the "word-shape" and linguistic experience render individual letter-testing too reductive?

A: Visual recognition of what? This is a valid methodology with a simple

convenience, that the shapes correspond to letterforms. You could be testing squares and triangles. So, you may be able to identify formal characteristics that are more reliably identifiable in isolation, but that has nothing to do with typeface design, where immersive reading takes you back to typographic parameters: apparent size (minutes of degrees), space above/below surrounding where your eyes fixate, visual semantics (e.g. segmentation, rhetorical devices, non-verbal cues and content) and linguistic / verbal content factors. Letters that perform well in pure recognition tests are, at best, a starting point for a scaffold towards readability and comprehension. So, yes, fine – as long as you do not see meaning in any results beyond what the framing of the experiment allows. (And as a reflective researcher, your first task is to be aware that your testing happens in discipline / funding / publishing / ... environments, which prescribe how anything is presented. In other words, most researchers do the work that their environment enables them to do, and have a career and esteem motivation to present it as meaningful. Treat this as education in experimental processes, and then learn as much as you can from trying to pick holes in your work, or find other ways of looking at the problem. Intellectual humility is not taught in universities, you have to learn it yourself.)

Q: You have pioneered the study of multiscript typography. My study is currently focused on the Latin script and Portuguese diacritics. In your experience, are the "legibility rules" for low vision universal across scripts (e.g., Greek, Cyrillic, Arabic), or do we need script-specific AI models to validate accessibility?

A: North-western approaches tend to assume that what works for their contexts (Latin, LTR, documents of certain structures, etc.) should work for other scripts, and it is up to those representing them to make the case for requiring script/language specific testing. This represents a biased view already. It is more logical to start from the assumption that a script/language/culture combination constructs different experiences, and to identify any similarities in behaviour on the basis of converging independent results. This means more work, but it recognises local researchers, and local conditions. If there's one thing that is by now a certainty in social sciences research, is that researchers cannot be trusted to audit their own assumptions and biases: you need to get external validation by people with an understanding of the context of work. This takes us back to the experience scaffold I mentioned earlier. If you are looking at fundamental optical matters, then sure, a Bengali reader with glaucoma is the same as a Kuwaiti, a Portuguese and a Chinese. But each will read differently, depending on what they read and how: are they reading textual patterns that are familiar? The worst sufferers from such assumptions are LTR alphabetic scripts, at the hands of Anglophone researchers. How many letters do you need to discern before the brain fixes on a word? And if you want to consider the shape or letterforms, how does the effect of visual resonance affect perception? (Arnold Wilkins at Essex did some interesting work looking at the frequency of letterforms with vertical strokes – Latin and Cyrillic – that create discomfort, whereas Greek with predominantly bowls has zero incidence of that effect.) So, yes: if you are going to assume anything, I think you need to start from the point of diverse performances as soon as you look higher than pure optical mechanics, and work to demonstrate commonalities. But I do recognise that your experiment is low in that scaffold. The same for any AI-driven models: what are the biases built into systems trained on primarily English-language outputs and test materials? I think there's a lot of potential for models with tightly curated training sets, but that is still expensive and time consuming.

Q: I think an opportunity for this research would be to use Variable Font axes to "tune" type for specific eye pathologies (Cataracts, Glaucoma). From a cur-

riculum and professional standpoint, do you think we are asking too much of type designers to become "visual engineers," or is this technical adaptability the future of the profession?

A: Yes, VFs provide a very good platform for exploring parameters in ways that have been too difficult before. (Octavio Pardo-Virto's use in his PhD is a good example.) The potential for VFs to adjust to viewer circumstances has also been demonstrated for AR/VR environments (see Niteesh Yadav's work). However, they also show the limitations of a font format that embodies assumptions about Latin-centric documents with very dated semantic structures, and a hard separation between font and text strings. (And TeX or Karow's work show that this was not a given even in the 70s and 80s.) But your second statement, about type designers, jumps through stages of discipline expertise and loads what are entry-level curricula with unrealistic expectations. Do you allow a civil engineer straight out of a BSc to build a bridge? Or do you scaffold their actions in a layered professional environment and protect with regulatory frameworks like accreditation? Typeface design has no Bachelor's level education, so people cram into short courses and Masters programmes a lot of learning. What is reasonable to expect people to learn in a year? (90 ECTS in Reading and ecal, similar in Amiens, ANRT, less so in Elisava and KABK, and so on.) If we had a more sustainable process, we would build fundamentals in a 3-year Bachelors, spanning design, technical skills, language fundamentals, cultural models, and research methods; then people could specialise at Masters level, taking paths – one of which could be empirically supported design. (Or a 4-year Bachelors, with a final year specialisation.) But the audience for such a specialist field at the point where people choose university courses, 17 or 18 years old, is negligible. This is a case where thinking narrowly about something ("is it good to train type designers as 'visual engineers'?") hits the problems of institutional models, funding, career paths, and so on. The problem is not one of answering the question in the absolute ("Yes, of course it is") but one of "once we decide that it is a good idea, what are the paths to realising this?" To be open with you, I've been thinking about this for over 20 years... one of my goals when I got involved with ATypI was to convert it into an accreditation body, so that we could provide a framework for this career portfolio. I think the only chance of anything happening is in a place like China, where you have three enabling factors: design happening in research-intensive universities; a very large potential market; and non-trivial complexity of the script. The third condition will be gradually diminished through AI tools, but there are training sample challenges, so the progress will be slower than in other fields. But it is still possible to imagine enough people interested to sustain a course, which is not possible in Europe.

Q: Designers often rely on their "eye" to judge if a font is clear enough. My research proposes to validate a tool that can accurately simulate legibility tests with visually impaired people. Do you worry that such tools might limit a student's creative intuition, or do you see them as a necessary "safety net" for inclusive design?

A: The designer's "eye" is a stand in for tacit knowledge – a good designer maps what they see onto received patterns. (What is "Regular" in terms of weight, and how does this evaluation shift over time? You know through experience and comparison.) This comparative mechanism – a contextualisation of judgement, if you like – can certainly be technically supported, and possibly substituted with technical means. AI tools could do this easily, if it is worth the effort to assemble and evaluate the training sets. (But a small field makes the returns on this questionable, so human labour has a competitive advantage.) I have a very low opinion of human "creative intuition" or "designers' eye" or anything that suggests a subjective "black box"

mechanism. In my experience these are all signifiers of sociocultural patterns and experiential motivations that we are not describing well. (Because we have built the discipline of design around the myth of the [usually male, older] singular genius, whose authority speaks for itself. That is constant in western and eastern cultures alike.) Tools as “safety nets” are a good idea, if you can build processes of external review and validation, and periodic re-checking – otherwise you are just building another mechanism of control, instead of a “revered designer”, a “regulatory check”.

Q: Assuming this toolset is accurate in the simulation of users with several types of visual impairments, this could change the way we design typefaces. How do we prevent these validation tools from stifling the typographic diversity and “flavor” that you have always championed?

A: There’s a leap from “the tool is accurate” to “this will change the way we design typefaces”. This is at least a category error, typeface design is primarily driven by commercial expediencies with a mix of functional requirements and formal qualities. “Adequate performance” is very much context-dependent. Readers with visual impairments (don’t call them “users”!) have their own context, which is technical, cultural, financial, and social. Let’s assume your toolset works, and you publish a couple of papers from it. So what? Is there a direct link from that to other teams validating your work (because surely you can’t expect a single set of experiments by one unit to be taken for granted), then adapting the results into practice recommendations? And who will fund the development and adoption of new typefaces (or run a campaign for statutory prescription)? Microsoft adopting a technology (like some of the ART outcomes, or the recent Kermit project) are exceptions, and tell you more about how R&D works in the tech sector and market capture than a shift in practice in a commercial market. I don’t think validation tools pose a risk to diversity of form any more than bad internal processes resulting in a regressive, unfit-for-purpose default family like Aptos, for example.

Q: You’ve been a leader in both academia and industry. What is the responsibility of the University in providing open-source validation tools like VI-OCR? Can academic research truly change the production standards of major foundries and tech giants?

A: You’re kind but I’d say that all I’ve learned is the result of failures to effect change, and anything lasting happens despite institutions, rather than because of them. Are you assuming that universities should do applied research? Some of it, for sure, but I’d prefer to see more speculative research in universities, and intentional breaking down of discipline boundaries. The whole point of research in universities is that it does not need to justify its outcomes or impact, but enable knowledge creation. That’s what a university is for – if you then make the knowledge (and tools, techniques, whatever) available to others, that’s the end of your “social contract” with the University. You are neither responsible for what will happen to your work, nor are you really to take credit for it – successful adoption of something is so much dependent on context that it is silly to draw causal relationships. (People do so, because it can be exploited, but that doesn’t mean much.) This doesn’t really include work at Masters level, this is really just training in doing, and hopefully thinking. And Masters work, like yours, is probably better suited for applied research, where the scale of investigation can fit within the time and resource constraints of a programme. But very worthy work at that level can go unacknowledged and that doesn’t mean it is not worth adopting commercially or through regulation. So, yes, academic research can change stuff, but you can’t know if it will, and you shouldn’t really be doing it anticipating the outcomes.

“Crowded” text is a major barrier for the visually impaired. In your teaching, you emphasize the “rhythm” of spacing. Do you think we can find a mathematical “sweet spot” for spacing that survives visual degradation, or is rhythm too dependent on the designer’s intent and the specific text?

This goes back to the context of “the act of reading”. Ignoring the roots of any script association with any patterns, this is definitely describable and “modelable” technically. (But of course “rhythm” only for Latin, Cyrillic, and scripts based on repeated strokes in the same orientation; for other scripts you need to talk about “texture” in two dimensions at least: what’s the expected pattern of relationships in Bengali, or Korean, for example? And in that case what “looks right” is very much dependent on the frequency of adjacent combinations.) Designers have much less agency in this than you suggest. They internalise the rhythm or pattern, and learn how to reproduce it in new designs. Good ones can push the envelope of what’s acceptable by readers, without distressing expectations too much. That’s it, really. But for visually impaired readers I’d be inclined to think more in terms of determining which parameters are worth giving them control over. They already control apparent size (either on screen, or by holding things closer); I’d expect a good starting solution to offer adjustments in tracking and weight. Experiments can help us understand the priorities for adjustments, that should then be made available to readers. We don’t want to replace the dictatorship of the print typographer with one by a designer with a couple of citations on their CV.

Q: You have previously emphasized the human-centric nature of the design process. In an era of AI, where tools like VI-OCR are being developed and tested, how do we ensure that the “design process” remains the core of the discipline? What would be an ethical way (if there is any) to propose an integration of these tools in the type design process?

A: Why do you think that new tools threaten in any way the “design process”? The “process” is not what gives value to outcomes in applied fields, the suitability and performance of the outcome for the specific purpose matters. (It is useful to compare with architecture: when, to what degree, and why do architects discuss “process”?) Specific designers may avoid new techniques, but that doesn’t mean anything. Inquisitive designers will explore the value of each new tool for the specific challenges they face, and adopt them if they improve the efficiency of their work. (This may be speed, cost, before it is quality, if history is any lesson.) Your tool may be instrumental in one category of design problems, and irrelevant in another. Designers will determine whether it is useful to them or not, and that’s fine. Your questions have one foot in the narrow space of doing an experiment as part of a Masters degree, and one in much broader questions about the discipline of typeface design, and its relationship to other fields; how knowledge is created, and what role, purpose and known outcomes have in this. That is a compliment to you as a person and reflective learner, but arguably very much outside the scope of your testing.

Erik Spiekermann

Q: When designing for the German Railways (DB), you had to account for poor lighting and motion. How do you “stress-test” the anatomy of your letters during the drawing stage to ensure they don't collapse into illegibility for someone with reduced contrast sensitivity or low visual acuity?

A: After 50 years of type design I know what to look for. I print out and look at screens all the time – in different sizes, positive and negative, in different colours and sizes – like every type designer should.

Q: This research uses VI-OCR to simulate several types of visual impairments. You've always relied on intuition and physical proofs. Would a tool that gave you real-time “Legibility Scores” or any other relevant data have helped you drawing your typefaces, or is there a human “gut feeling” about legibility that an algorithm simply cannot replicate?

A: There is more than a gut feeling. When I had finished the first version of PT55, the typeface for the German Post Office that later became FF Meta, I sent the first version to Heiner Erke at Braunschweig University, who was Professor for Wahrnehmungspsychologie – cognitive psychology – who tested the legibility. I don't know his methods but he told me that certain details in the typeface were a little too exaggerated and cause noise. If you look at FF Meta, you can still see some of those details, but they were more explicit in the first version. When I had designed TERN for the European Motorways, the type was tested at Budapest University. Unfortunately, the method was quite primitive: a group of people in a dark room were shown single letters on a screen and it was measured how long it took them to recognize each letter. In real life, however, nobody reads single letters. Some of the remarks made sense while most of them simply confirmed what typefaces people were familiar with. I'd love a proper tool that took into account people's cultural background, not only physical details.

Q: You've said that “type is a service”. For the 250 million people with visual impairments, a font is a survival tool. At what point does “design personality” have to give way to ensure legibility, and have you ever had to “kill” a design feature because it failed the service of reading?

A: Every letter in our alphabet (considering only “serious” type for reading complex messages, long text, signs, etc – not advertising or packages) has to look like every other letter with the same meaning – an “a” has to look like an a, within a very small space of variations. As type designer, I work within a space of some 10% that a letter in one typeface can be different from that same letter in another face. It is often enough to just make a few letters different to distinguish a typeface from another. I can identify most typefaces by looking at the lower case a. The g, e, and perhaps k will tell me a little more, but the rest of them have to pretty much follow a well-known pattern – depending on the audience's culture, age and background.

Q: A core part of my research is mapping “Character Confusion”, where letters like 'a' and 'o' or 'l' and '1' merge for the visually impaired. You've always fought for “distinctive” character design. Why do you think the design world is still so obsessed with “neutrality” and “minimalism” when those choices are so exclusionary to people who don't have perfect vision?

A: Most graphic designers do not read very much and consider the effect of a page, not the legibility of a word or line. They look at type at making a pleasant pattern. What works for visually impaired people works especially well for “normal” readers because everybody gets into situations that make reading difficult.

Q: Variable fonts allow a user, for example, to change the weight and width of a font at will. As someone who has spent decades perfecting the “static” weight of a typeface for maximum performance, does the idea of “user-adjusted” typography feel like progress to you, or does it feel like a surrender of the designer’s responsibility?

A: Only if the user gets led through the process. Nobody knows what width, tracking, kerning or other typographic parameters mean, but if an interface lets them play with the type until it looks good to them, then this would be a good tool.

Q: Most type is designed for the “standard” eye. What helps a Glaucoma patient might actually hurt a Cataract patient. In a digital world, should we stop trying to design “Universal” fonts and start designing “Adaptive” ones?

A: The Universal has to be good before we should make it adaptive. A bad typeface doesn’t get better by changing some parameters. Besides, other features like line-lengths, tracking, line-space (formerly leading) are often much more important than the choice of type. The best typeface is useless when there are 120 characters per line or when the type is set too tightly.

Q: If you could set one “Grand Challenge” for the next generation of designers regarding visual impairments and the future of reading, what would it be?

A: Read more yourselves, study thousands of typefaces (not just the trendy ones), learn about the science of reading as well as the science of understanding and how our visual apparatus as well as our brain works.

Q: Usually, only big companies can afford clinical legibility studies. If my research proves that a free, open-source toolset can accurately simulate low-vision failure, does that empower the independent designer, or does it lead to a world of “safe” but boring typography?

A) Typography is not boring. B) I’d love a toolset like this but most of my fellow graphic designers and even typedesigners would be too lazy or too arrogant to let science interfere with their “artistic” intuition. I am a designer, not an artist, and my job is to make visual communication easy to understand and pleasant to behold.

Q: Helvetica is the “default” for many, yet it is often cited as a nightmare for low-vision readers due to its closed apertures, and many other flaws. Why does the design world continue to celebrate “neutrality” when it is so exclusionary to a large portion of the population?

A: We’re all lazy, afraid to take risks and it is easier to follow examples and trends than to fight the mainstream. Nobody has ever been blamed for using Helvetica. Or for eating white bread.

Q: You’ve seen type move from metal to film to pixels to now AI. At the end of the day, a visually impaired reader just wants to get to the right train platform. Is the “tech” (AI, OCR, Simulations) getting in the way of us simply being better, more empathetic craftsmen?

A: As long as we know how to use these tools, they can help us. A real craftsman uses his/her tools until they become second nature. From then on, you can relax and make good things without having to be afraid of making mistakes.

Aprígio Morgado

Q: At ESAD.cr, you work with students in a very experimental and hands-on environment. Do you think integrating "Legibility Simulation" tools like VI-OCR into the curriculum could shift the student's perspective from "aesthetic-first" design to a more "empathy-first" or "evidence-based" methodology?

A: Yes, I think that the introduction of user-friendly legibility testing tools would allow for discussions that are less focused on graphic languages and more focused on their relationship with content and contemporary societal issues.

Q: You have a deep understanding of communication design in the public sphere. When a typeface is used in the "wild", on signage or public information, it faces environmental stress (blur, distance, low light). Can a computational model ever truly capture the "chaos" of how we read in a real city?

A: Not at the present. The reading environment in the city depends on many uncontrollable factors. Perhaps in the future there will be models that can do this.

Q: Designers often rely on their "eye" to judge if a font is clear enough. My research proposes to validate a tool that can accurately simulate legibility tests with visually impaired people. Do you worry that such tools might limit a student's creative intuition, or do you see them as a necessary "safety net" for inclusive design?

A: In my view, the academic discourse of graphic design courses remains overly focused on graphic languages. I believe this is partly due to the fact that graphic design has historically shared the tools of the visual arts. This legacy has made us confused professionals, unaware of the transformative potential of design. Therefore, I do not see these tools as a limitation, but rather as an opportunity to question institutionalised design practices, focusing on concrete problems in society. For example, according to INE estimates, by 2050, one third of the Portuguese population will be over 65 years old, which means that a large part of this population will have visual acuity problems.

Q: Many designers may see accessibility as a "burden" that limits their style. Could we reframe accessibility as a "Creative Constraint", similar to the ones we use in design exercises, where the goal is to survive a specific "Blur Threshold" or "Character Confusion" test?

A: Framing accessibility as a creative constraint that limits visual style assumes that the designer's creative freedom takes precedence over the social relevance of an accessibility project. I don't see it that way; I think that is a very postmodern view of design, where questions of authorship, subjectivity and self-promotion prevail.

Q: I am using Kevin Larson's 17ms trigram protocol to test how fast an AI recognizes characters. In communication design, we often talk about "glanceability" (e.g., seeing a billboard or a bus number). Does this split-second recognition matter more for inclusion than the "long-form" reading experience?

A: I think it depends on the type of visual impairment and its severity. Testing 17 ms trigrams may be inappropriate for some conditions, and the same pathology can present quite differently from one person to another. One of the design projects I find most compelling in the field of inclusion is Displaying Dyslexia, by designer Sofie Gagelmans. She studied dyslexia and recognised how differently it can manifest across individuals. So, rather

than attempting to produce a one-size-fits-all solution, she developed a console that helps simulate and communicate – especially to people without dyslexia – the visual symptoms a person with dyslexia experiences, enabling a clearer understanding of the condition.

Q: Variable fonts allow for infinite adjustments in weight and width. Do you see a future in Portugal where public digital services (like Citizen Shops or Health Portals) automatically adjust their typography to the user's visual profile, or is that still a "technological utopia"?

A: I am confident that it will. I believe that variable fonts open up new possibilities for inclusive design and, in this sense, I believe that research on the subject will be the first step towards the eventual implementation of visual interfaces with this technology. Its implementation will depend on scientific evidence and political will. In Portugal, what we know is that the norms and legislation applicable to public digital services are the result of European directives.

Q: What do you believe is the responsibility of design schools like ESAD.cr and FBAUP in developing or promoting open-source tools for the community? Does a tool like VI-OCR fulfill the academic mission to bridge the gap between high-level research and the local design market?

A: I like the idea of a school that promotes community-oriented design practices by providing tools that challenge the market's dominant logic. I believe that making a tool such as VI-OCR available would help to foster socially responsible design practices and, in that sense, it could be adopted by small studios with a strong design culture. However, I see the gap between high-level research and the market as a problem rooted primarily in business culture, which continues to rely on outdated paradigms of production.

Q: We often default to "safe" fonts like Arial or Helvetica for public information, yet research shows they fail under low-vision conditions. Why is the design community so attached to these "neutral" typefaces, and how do we break this cycle in the next generation of designers?

A: Studies show that younger generations prefer sans serif fonts, probably because they have been exposed to reading on digital devices from an early age. Initially, low-density screens favoured the simple shapes of sans serif fonts for reasons of legibility, which may explain part of this predisposition. The adoption of so-called neutral fonts may be due to market trends. In recent years, there has been an adoption of grid-based languages, whose rational structure easily adapts to the logic of templates and UI engineering/development. The way in which the grotesque letter reinforces the line, the block of text and adapts to the grid is, I believe, the reason for its resistance to time. On the other hand, branding seems to perpetuate a eurocentric corporate visual culture, where the legacy of international style is still present.

Q: Clinical trials with visually impaired participants are hard to organize for independent designers. If my research proves that an AI simulation tool can accurately "stand in" for a human reader, do you think this empowers the individual designer to compete with large agencies that have big research budgets?

A: Yes, if the tool remains open source.

Q: You have seen design evolve through many technological shifts. Is there a danger that by focusing so much on "legibility scores" and "OCR accuracy," we forget the "pleasure" of reading or the emotional connection a user has with a typeface?

A: Yes, this is always the problem if we assume that font selection should always follow legibility efficiency criteria. A font can be legible effective without being efficient – the need for efficiency depends on the context and purpose. Anyway, studying emotional factors that influence reading quality is different from studying legibility. Since the beginning of the last century, there has been a need to distinguish between character recognition (related to low-level visual processing) and continuous reading performance, which is influenced by factors such as language, attention and context.

James Edmondson

Note: This is a transcription from an audio file James kindly sent me via email.

Q: You often talk about "flavor" in type, the soul and personality that makes a font unique. In my research, I've found that high-personality shapes can sometimes provide better "anchors" for a struggling eye than "neutral" fonts. Do you think a font can have "too much personality" for a low-vision reader, or is the "invisible type" myth actually harming accessibility?

A: I don't know. I have no way of testing this. I only have hunches. So my answers to probably most of these questions are going to be guesses at best. I think at the end of the day, we read best what we read the most. And it doesn't matter whether what we read most is something closer to a neutral looking font or a really expressive font. But it all depends on the reader's background. So if the reader has spent their lifetime reading something that is extremely expressive or even something just like handwriting, then that's the thing that they will read the best. I don't think it, I don't think our eyes really have too many biases towards specific forms. I think it is more about how we were taught and what we're comfortable with.

Q: As a designer who relies heavily on intuition, would you find it useful to have a computer tell you a character needs improvement from the perspective of accessibility of visually impaired users or do you think that kind of data stifles the creative risk-taking that Ohno is known for?

A: In my experience, I'm using the computer sometimes as a second opinion. If a computer was to tell me that the glyph that I'm working on is at risk of being illegible, I would sort of take it with a grain of salt. I think it is extremely easy for people to create legible typefaces. We've done it thousands and thousands of times. So that's not really the question that I'm trying to solve with a lot of my work. The thing that I'm more interested in is how do we do that while doing something else, like while making a visual impact or while conveying some sort of interesting emotion or feeling? Yeah. So I would be very skeptical on what data the computer was trained on for knowing what is legible and what is not legible.

Because like I said, everyone's legibility preferences are different. There is no unified thought of what is legible and what is not.

Q: Ohno designs a lot of incredible display type. In your mind, is there a difference between the way you develop a typeface for the reader of a display font versus a text font when it comes to accessibility? Should display type be "allowed" to fail sooner?

A: Yeah i mean obviously you can play more with legibility/illegibility in display type and text type is just smaller, it is the same problems we're trying to solve but they're just getting rendered at smaller sizes so i don't see too much of a design of a distinction in in how i'm drawing them, you know, i'm still drawing them the same way but i'm just evaluating them at different sizes not not too different really.

Q: When you are drawing "weird" or expressive shapes, how do you ensure that the letterforms don't just become "noise" for someone with low visual acuity? At Ohno what kind of methods do you use to test the legibility of your typefaces?

A: We do not use any method to test the legibility of our typefaces. It is sort of like asking a musician, how are you writing music for someone with low hearing? And really, I am not thinking of people with visual impairments

when I'm working on my typefaces. For the most part, again, I'm thinking about the size that they're going to be rendered.

Q: Variable Fonts allow for infinite weights. I think using these axes to "tune" type for specific eye pathologies (like Cataracts) can offer a lot of possibilities. Do you see Variable technology as a way to make "fun" type accessible, or is it just more engineering work for the designer?

A: I've not thought about variable technology in that way, but I do think that the potential uses for variable technology are totally untapped. And that is maybe one avenue of many that have not really been explored fully.

Q: Clinical trials are expensive and usually time consuming. If an open-source AI (VI-OCR) toolset can accurately mimic a low-vision reader, do you think this helps smaller independent foundries run more legibility tests for their new releases?

A: Yeah i think that's a huge "if an open source AI, can accurately mimic a low vision reader?" What type of low vision reader? Yeah, that would be tremendous if it indeed could. Again, I would be extremely skeptical of an AI tool doing something like that.

Q: By contributing with the development of an open-source tool, we could say there is an attempt to democratize testing. What is, in your view, the ethical responsibility of a designer/researcher when developing tools that might eventually dictate accessibility standards?

A: Again, this is a question I've literally never thought about before. The ethical responsibility of a designer is something that is up to them. And it is against my own ethics to define what other people's ethics should be. So, yeah, maybe that's a way of just getting around that question.

Q: The design world is obsessed with "safe" sans-serifs. Why do you think we are so attached to these "neutral" faces for "serious" information, and how can a foundry like Ohno help convince people that "distinctive" is actually "safer" for a reader who can't see well?

A: I don't think I can convince people of that. I think it would be sort of a fool's errand. Again, I think we just read best what we read the most. Over the last handful of decades, we've seen a switch from people reading mostly long form stuff in books to that printed mainly in sans serif or in serif text fonts at like eight to 11 point, and now we've seen a switch to people like I'm reading this PDF in a switch to mostly sans serifs, either because screens render them with less trouble or just because they're the trend of the day. So much of this stuff is sort of arbitrary at the end of the day. What they were reading in 15th century Germany was to them the most legible. And if we try to read that now, it is very difficult.

Q: At the end of the day, typography is a human-to-human connection. As we move into an era of "AI-everything", how do we make sure we don't end up with a world of perfect but totally boring fonts? How do you feel about the most recent developments in AI in the type design industry?

A: What we use AI for mostly is helping us write emails. I don't think we need to use it for anything. I think there will always be a place for expressive typography. There will always be artists, designers, craftspeople who want to make it. And yeah, I do not think that is a threat at all.

Q. As a designer, you are usually the "owner" of the curve. However, when we talk about low vision, the features we design often become components that affect how information is perceived. If a "personality-driven" flourish in your design causes an OCR or a human to misread a medicine bottle or an emergency sign, where does your creative freedom end and your ethical responsibility to the reader begin? How can we balance creating unique, creative typefaces and ensuring readability?

A: I think you might be giving me a little bit more credit than I deserve here. I wake up in the morning with an urge to draw and urge to create, um, I am thinking how to explore the inner recesses of my mind and my creativity. And I'm not thinking about the user. What I love is that the user gets to look at my work and the work of all of my colleagues, past and present, and decide for themselves what is going to work for their project. So these are questions that are largely not up to me. I admit that's a kind of luxurious place to exist where we create what we want, we release it into the world and we see what happens. I've worked pretty hard to get to that point. That's always been my dream is to work in that way. And yeah, I'm happy to be for the most part around there. Some days are obviously more work than others, but yeah, I'm not thinking about an ethical responsibility with my work. And I do think about the reader, but only as much as I am reading what I'm looking at. I can't really imagine what it is like for someone else to read things. I can guess, but I'll never really know. I don't know how other people see color or I don't know how other people feel texture or experience music or all this other stuff. You'll never get in someone else's perception. So in my mind, it is kind of a waste of time. What I like to do is create things that I haven't created before. I even like creating things that I have kind of created before. I think it is just the physical act of drawing and spacing is what I enjoy. And to do that takes a complete relaxation or like a turning off of the part of my mind that's an editor or a skeptic or someone with an ethical responsibility. Like when I'm drawing, I'm only a creator. And then I can step back from it and say like, okay, does this really have the legs to move on further? And most of the time it doesn't.

Dan Rhatigan

Dan Rhatigan Chat Via Zoom 15th May 2026 regarding font distribution, licensing and other topics

Note: This conversation was in a very informal tone as me and Dan already knew each other fairly well from hanging out at the 2026 Fontstand Conference in Berlin. A few parts of the conversation were removed as per Dan's request.

This transcription was also edited to remove weird pauses or certain language specific mistakes, in order to help build a more coherent conversation easily understandable by the reader.

Transcription

[00:00] Hugo Carvalho: ... sure.

[00:01] Dan Rhatigan: Record. There you go.

[00:02] Hugo Carvalho: Okay.

[00:02] Dan Rhatigan: Just like that.

[00:03] Hugo Carvalho: Cool. Yeah.

[00:04] Dan Rhatigan: Yeah.

[00:05] Hugo Carvalho: So, this can be very informal. (laughs) you know? Although we are recording, this can be very informal.

[00:12] Dan Rhatigan: Yeah.

[00:13] Hugo Carvalho: First of all, again, thank you very much for doing this. It is very cool.

[00:19] Dan Rhatigan: It is cool.

[00:19] Hugo Carvalho: You are one of the people that I met at Fontstand, that I was so excited to talk with again for many reasons, mostly because you have a lot of experience, but also because I think you are just a cool guy. So... (laughs)

[00:35] Dan Rhatigan: (laughs) You don't have to butter me up.

[00:37] Hugo Carvalho: No, i- it is not about buttering.

[00:39] Dan Rhatigan: (laughs)

[00:39] Hugo Carvalho: It is about, saying the stuff that's true, you know?

[00:45] Dan Rhatigan: (laughs) Right.

[00:45] Hugo Carvalho: I'm also a bit of a fan, of course, so it ends up lining up with that. (laughs) Like, I was talking with Octavio Pardo, just now through LinkedIn, sending him a few messages, and I was like, man, You reach a point that you are like, "It is so cool that I can have all these conversations," but I'm like a bit of a fan, and I'm like, "Oh, man." (laughs)

[01:07] Dan Rhatigan: (laughs)

[01:08] Hugo Carvalho: "Where do I go from there?" Um, so you get a little bit lost, but yeah, um, it is very cool to, like, be able to say stuff like, "Oh, yeah, by the way," which is, is going to happen now that I'm talking to you, like, adding parts to the dissertation where I say, "By the way, I had a conversation with Dan Rhatigan about this."

[01:26] Dan Rhatigan: (laughs)

[01:26] Hugo Carvalho: "And he told me this, and I thought this was interesting." Like, that's cool.

[01:29] Dan Rhatigan: Mm-hmm.

[01:29] Hugo Carvalho: Like, saying, for example, I have a part on my dissertation where, um, I talk about, several different topics, and I'm like, "Oh, by the way, Dave Crossland introduced me to Octavio Pardo, and Octavio told me this, and I found his work-"

[01:44] Dan Rhatigan: Yeah.

[01:44] Hugo Carvalho: "... very exciting," and stuff like that. So that's like... It is a bit of a flex for me, kind of-

[01:49] Dan Rhatigan: (laughs)

[01:49] Hugo Carvalho: ... but it is a cool thing to say. It is a very nerd... If you are into type design, it is a very cool thing because it is almost-

[01:56] Dan Rhatigan: Yeah.

[01:56] Hugo Carvalho: ... like you are calling out all these cool people-

[01:58] Dan Rhatigan: I know.

[01:58] Hugo Carvalho: ... that you end up meeting.

[01:58] Dan Rhatigan: I've been there. I know what you're talking about. Yeah.

[02:01] Hugo Carvalho: Oh, man. It is so cool. So, how are you doing? (laughs)

[02:06] Dan Rhatigan: I'm doing all right.

[02:09] Hugo Carvalho: You are back in Berlin, right?

[02:10] Dan Rhatigan: I'm back, I've been back in Berlin for about three weeks. I'm gonna be here for a week.

[02:14] Hugo Carvalho: Ah, that's cool.

[02:15] Dan Rhatigan: And I've got another week, then I'm gonna go to... Well, I'm now going to Zurich for a day, on Friday, for a meeting.

[02:22] Hugo Carvalho: Oh. That's really cool.

[02:23] Dan Rhatigan: And then the next day, I'm going to Vienna for a little over a week.

[02:28] Hugo Carvalho: And that's because of, like, Monotype stuff?

[02:30] Dan Rhatigan: Um, no. Vienna, 'cause I'm just, like... Monotype is not paying for me to be here. I'm just sort of working from their office here and hanging out in Berlin.

[02:39] Hugo Carvalho: Oh.

[02:40] Dan Rhatigan: But then, um, a good friend of mine... Do you know the Typejockeys?

[02:48] Hugo Carvalho: Yeah.

[02:49] Dan Rhatigan: Yeah. So, yeah. Mihi from Type Jockeys is a good friend of mine. We were housemates in Reading.

[02:54] Hugo Carvalho: Oh.

[02:54] Dan Rhatigan: So I'm gonna go stay with him and hang out with his family, and then just head back home for 10 days to essentially do laundry and swap out my bed.

[03:03] Hugo Carvalho: (laughs)

[03:04] Dan Rhatigan: Then I go to Barcelona, and then on to the -.

[03:07] Hugo Carvalho: And then Porto. (laughs)

[03:09] Dan Rhatigan: I have no, no side quests planned while I'm on the Iberian Peninsula, unfortunately. I don't really have time. 'Cause I have to get from Barcelona to Italy to teach for two weeks.

[03:19] Hugo Carvalho: Ah, that's cool. I mean, uh-

[03:21] Dan Rhatigan: It is, it is a lot.

[03:23] Hugo Carvalho: You're such a rockstar, I'll say this.

[03:25] Dan Rhatigan: (laughs)

[03:25] Hugo Carvalho: You're such a rockstar. It is very cool. Like, if people ask, and I'm not buttering you up, but-

[03:31] Dan Rhatigan: (laughs)

[03:31] Hugo Carvalho: ... if people ask me, like, "Hugo, who do you wanna be when you grow up?" I would say, "Dan Rhatigan." It is such a cool thing, what you're doing.

[03:38] Dan Rhatigan: Ooh, Monkey's Paw is gonna close a finger on that on you someday. (laughs)

[03:42] Hugo Carvalho: (laughs)

[03:43] Dan Rhatigan: Um, but let's get to your dissertation.

[03:45] Hugo Carvalho: Maybe, maybe one day, when, when I write a book, I get to write the introduction for the book, and you can call this up, this meeting up.

[03:51] Dan Rhatigan: Yeah, that'll work.

[03:52] Hugo Carvalho: (laughs) Uh, yeah.

[03:53] Dan Rhatigan: I don't, I don't want to waste too much of the time 'cause-

[03:55] Hugo Carvalho: Yeah. Sure. Uh-

[03:56] Dan Rhatigan: ... there's perhaps a lot of information I can give.

[03:58] Hugo Carvalho: Getting to the point, yeah.

[04:00] Dan Rhatigan: Yeah. What do you wanna know?

[04:02] Hugo Carvalho: So, um... Okay. Um, what I was writing about, th- the reason I asked this question was because I was writing about, uh, distribution publication. I was essentially kind of breaking down, uh, kind of the state of the type industry in general, type design industry-

[04:20] Dan Rhatigan: Yeah.

[04:20] Hugo Carvalho: ... in general. And I was, uh, I have a point where I talk about distribution and publication, all of that. And I essentially want to go through every way that you could potentially commercialize, typefaces. And I talk about stuff like marketplaces, large marketplaces, such as-

[04:39] Dan Rhatigan: Mm-hmm.

[04:40] Hugo Carvalho: ... MyFonts and stuff like that, um, or Fontstand or whatever. But also about alternative ways that have been coming up, like, uh, mostly relating to licensing and stuff like that, like subscription-based and stuff like that.

[04:53] Dan Rhatigan: Yeah.

[04:53] Hugo Carvalho: I call out Fontstand 2026, in Berlin, that we went to. And I call out how they proposed a new model and how that could be interesting.

[05:01] Dan Rhatigan: Yeah.

[05:01] Hugo Carvalho: Um, and also, other stuff like, uh, open source stuff, essentially. So, what I wanted to know was essentially opinions of, in this case, you, on, you know... What do you think are the main ways for distribution, distribution of typefaces and-

[05:23] Dan Rhatigan: Mm-hmm.

[05:23] Hugo Carvalho: ... commercialization of typefaces. Um, and also, your opinion on marketplaces or self-publishing and all of that. Because in general, when you talk to people, it is kind of a gray area. Um...

[05:39] Dan Rhatigan: Yeah.

[05:39] Hugo Carvalho: But yeah, that, that's pretty much it.

[05:42] Dan Rhatigan: Yeah.

[05:42] Hugo Carvalho: Uh, and also, one thing that I, I did not find much information about, but I think it would be cool is about marketing your catalog. If you have-

[05:50] Dan Rhatigan: Mm-hmm.

[05:50] Hugo Carvalho: ... like, any information on that. In what ways do you, for example, reach designers to sell fonts? I think that, that's also a very cool thing if you could talk about that.

[06:02] Dan Rhatigan: That's what everyone's concerned about now 'cause it has shifted very dramatically.

[06:06] Hugo Carvalho: Mm-hmm.

[06:06] Dan Rhatigan: Um, so...Okay. At the core of this ... would be ... if you've designed a typeface, um, one step ... The first step is gonna be sort of like how it is gonna get published.

[06:26] Hugo Carvalho: Mm-hmm.

[06:26] Dan Rhatigan: Distribution sort of extends outward from there. So, one, you could release it under an open source license, which means you just, it needs to become publicly available under that license. So, depen- up to you, like, do you host it on GitHub? Do you contribute it to Google Fonts? Um, but open source essentially means you released it, even if you monitor the projects, and you kind of let it go for free.

[06:59] Hugo Carvalho: Yeah.

[07:00] Dan Rhatigan: Um, you could, um, have someone else publish your typeface, so sign a deal with a foundry or something where, you know, you turn over the design, probably get royalties, occasionally maybe a buyout, but usually some kind of a royalty arrangement in order for another foundry or company to take on the responsibility of, like, finishing the font, handling sales, handling marketing, all of that stuff.

[07:31] Hugo Carvalho: Yeah.

[07:31] Dan Rhatigan: Or, you could choose to publish yourself. Start a f- start a foundry, well, start a fou- a company as a foundry, um, or, you know, even just publish it as a sole designer and make it available for sale in various ways. Um, but I just think those are, each of those are very distinct starting points, because they, they speak a lot to, like, what are the options for distribution in the future of the typeface? So,-

[08:03] Hugo Carvalho: Mm.

[08:04] Dan Rhatigan: ... open source means you're committing to, you're not gonna have any revenue from that typeface. You could only ... the return on that investment, you know, your time that you put into it, would just come from, you know, the exposure you get from it that leads to other opportunities. Like say-

[08:21] Hugo Carvalho: Yeah.

[08:21] Dan Rhatigan: ... maybe you get a later commission to do some, you know, specific additions to it for a fee. Again, not royalties, but for, like, a design fee. Or it just helps people know who you are and other work comes out of that another way.

[08:34] Hugo Carvalho: Yeah, tha- that's what I, I got from talking to, for example, Dave Crossland-

[08:38] Dan Rhatigan: Yeah.

[08:38] Hugo Carvalho: ... and stuff like that, and the Google Fonts model-

[08:41] Dan Rhatigan: Yeah.

[08:41] Hugo Carvalho: ... um, and how they, they approach that as well.

[08:43] Dan Rhatigan: Yeah. And so, the ... and, and a lot of the Google Fonts model does specifically, you know, they encourage people to, like, commission designers who contributed things to the project. They were like-

[08:57] Hugo Carvalho: Yeah.

[08:57] Dan Rhatigan: ... Google doesn't support those fonts. They encourage, you know, they try to connect people to make those things happen if they need any more work. So, if you publish with a foundry, which I think for a long time that was considered, like, the more likely route to go for newer designers. Like, not ... You know, the easiest way to sort of, like, enter the landscape, learn the rest of how it goes. Um, I've seen that change a lot in the last few years. Um, now that, like, I see the scale of the sheer number of people who release things, say, to MyFonts/Monotype ...

[09:38] Dan Rhatigan: Um-

[09:38] Hugo Carvalho: Yeah.

[09:39] Dan Rhatigan: ... the ... It is easier than ever to not have to have someone else publish your fonts. Um, it, the marketplace has gotten very dispersed. Whereas, you know, like, almost 20 years ago when I was finishing Reading, I was looking for someone to publish my typeface, 'cause there was a lot I did not know about the industry. It seemed like there were just a few major, like, marketing and distribution powerhouses. Um, so, you know, the idea there would be you make a proposal to someone. They help you finish it. They help you get it out into the world.

[10:20] Dan Rhatigan: And depending on the contractual arrangements, or either, you either share ownership of the fonts, or maybe you retain the ownership, but it is, like, very much under contract-

[10:30] Hugo Carvalho: You get royalties.

[10:31] Dan Rhatigan: ... per s- but someone else calls the shots besides you.

[10:34] Hugo Carvalho: Mm-hmm. Yeah.

[10:35] Dan Rhatigan: That's, that's sort of what you agree to, is that your work becomes part of someone else's product portfolio. Um, and, you know, there's still a lot of people who do that. Um, TypeTogether does that. You know, they publish, um, you know, they give a prize to help people publish a font.

[10:53] Hugo Carvalho: Yeah.

[10:53] Dan Rhatigan: They publish fonts from other people. Um, so-

[10:57] Hugo Carvalho: They, uh, do the, uh, Gerard Unger Scholarship.

[11:00] Dan Rhatigan: Yeah. Um, and you'll still find other, other people do-

ing that, and it tends to often come out of people who are, like, looking for students doing interesting new work, or it is a way of people supporting, like, designers they know who they've been freelancing with, so they get to know what they do, so it w- you know, a relationship grows out of that. Um, if you are going to publish your own fonts, the distribution question becomes something you have to think about more.

[11:34] Dan Rhatigan: 'Cause again, if you are s- if someone else is publishing your typeface,-

[11:38] Hugo Carvalho: Mm-hmm.

[11:38] Dan Rhatigan: ... you probably have no say into what their distribution plans are. It is prob-

[11:45] Hugo Carvalho: Yeah.

[11:45] Dan Rhatigan: ... you know, you're probably, the contract you have probably says what they, what they think is best for their business is what they can do. Um, you know, there may be prohibitions against them, like, you know, selling off your typeface. Um-

[11:59] Hugo Carvalho: Hm.

[11:59] Dan Rhatigan: ... but if you're gonna publish yourself, so I know ... So I'm gonna speak a little bit about Bijou as an example, because-

[12:06] Hugo Carvalho: Yeah.

[12:07] Dan Rhatigan: ... I've confronted a lot of these choices and have explored a lot of the different avenues. So the, um ... Actually, before I get to that, there is one kind of like a medium route, which I don't know if anyone else has tried it yet, but, um, it is the, uh, the Future Fonts route.

[12:28] Hugo Carvalho: Yeah.

[12:28] Dan Rhatigan: Where basically you sort of publish through Future Fonts, they have a stake in it, but it is like up to a certain point. You get to a point where like they say a family has graduated, in which case the designer can sort of have it back. So they're like temporary custodians of the font.

[12:48] Hugo Carvalho: Mm-hmm.

[12:48] Dan Rhatigan: I don't think... I don't know of anyone else yet who's following quite that same model. Um, so, you know, I don't know if it... You know, it would be interesting. Travis and Lizzie are lovely. Um, if you do want to talk to them more about that, if you think it is a significant enough thing to weave into this, I'm happy to write an intro... an email introducing you to them.

[13:11] Hugo Carvalho: I would love to talk to them.

[13:13] Dan Rhatigan: Yeah.

[13:13] Hugo Carvalho: It would be cool.

[13:13] Dan Rhatigan: Um, they're delightful. Um-

[13:16] Hugo Carvalho: I would imagine so.

[13:17] Dan Rhatigan: Yeah.

[13:18] Hugo Carvalho: They, seem very... The Futures Fonts model and the way Future Fonts is made seems very human, which is very-

[13:26] Dan Rhatigan: Yeah.

[13:26] Hugo Carvalho: ... interesting from an outsider's perspective.

[13:29] Dan Rhatigan: Yeah.

[13:30] Hugo Carvalho: It is very interesting. Seems very human.

[13:32] Dan Rhatigan: Yeah. Yeah. And people from Portland are cool, so, I mean, what can I say?

[13:35] Hugo Carvalho: (laughs)

[13:36] Dan Rhatigan: Um, but so, so if you're going to go the route to publish-

[13:39] Hugo Carvalho: Yeah.

[13:40] Dan Rhatigan: ... your own fonts, a lot of these things work the same way whether you form a company around what you're publishing or if you just are like, you know, an individual designer setting things out. So you, you always have the option of setting up a website making a font available for sale. That can go, you know, like however you set up the payment structure work, however you set up the website to work. Ultimately, you are taking responsibility for taking money and issuing a font with some kind of a license.

[14:19] Hugo Carvalho: Yeah.

[14:19] Dan Rhatigan: And there are like total- totally many levels of complexity to that. Um, but it also means that like you really are responsible then for everything that comes your way. Um, marketing, sales, contracts, special requests, um, all of it. Um-

[14:38] Hugo Carvalho: Yeah.

[14:38] Dan Rhatigan: So the easiest path, um, is to release a font and just turn it over to some distribution network who does like the reselling in exchange for some kind of royalties. So that could be Creative Market, it could be MyFonts/ Monotype Fonts. Um, I don't... You work for them, I don't know if they just sort of like take blind submissions or they just work stuff from a lot of people-

[15:13] Hugo Carvalho: Mm-hmm.

[15:13] Dan Rhatigan: ... or have a look closely at it. But basically, you're essentially saying it is like, "Okay, I've made a font. I don't want to deal with the business of selling it. So I'm gonna place it on someone else's e-commerce platform or something and-"

[15:28] Hugo Carvalho: And see where it goes.

[15:29] Dan Rhatigan: And just see where it goes. Like I said-

[15:31] Hugo Carvalho: Mm-hmm.

[15:31] Dan Rhatigan: ... this is where like Future Fonts is a bit of a hybrid,

because they're sort of like working with you as you go along.

[15:39] Hugo Carvalho: Yeah.

[15:39] Dan Rhatigan: Um, and they don't, they don't just take anything that people give them. You sort of like propose that you join the platform. But for, um, Creative Market or MyFonts, you upload a font, you like click to sign through an agreement, and then they sell it, you get royalties and ideally some stats about how what gets sold. Um, but, you know, the nature of royalties is you're getting paid for your intellectual property.

[16:12] Dan Rhatigan: Someone else is getting paid for all the effort-

[16:14] Hugo Carvalho: Yeah.

[16:14] Dan Rhatigan: ... that goes into getting yours out in the world. Um, there's a whole bunch of distribution models, and this isn't even getting to licensing models, but there's a whole bunch of distribution models to think about because, you know, part of choosing a distributor is getting on board with how they sell fonts. Um, you know, what kind of context are you gonna be sitting in? What sort of licenses do they sell? What are the terms of engagement, um, for you for working with them? So, uh, when I started Bijou, um, I began making fonts after I left Adobe because for the first time since I'd started working, I was no longer working for... full-time for a company that would claim ownership of anything I designed.

[17:09] Dan Rhatigan: That was-

[17:09] Hugo Carvalho: Mm-hmm.

[17:09] Dan Rhatigan: ... the case when I was at Monotype originally. It was the case at Adobe. So I just sort of sat on ideas until I could produce them on my own and make them available. Um-

[17:20] Hugo Carvalho: Cool.

[17:21] Dan Rhatigan: And then the, um... I did not... I was very, very slow about dealing with the idea of selling it myself. I knew like-

[17:32] Hugo Carvalho: Mm-hmm.

[17:32] Dan Rhatigan: ... "Ugh, I need a website and I need to think about what a EULA is, and I would need like an e-commerce system." So-

[17:39] Hugo Carvalho: There was a bit of a resistance there?

[17:42] Dan Rhatigan: Well, not a, a resistance, but it was just like, "Ugh, the to-do list is getting like too big."

[17:47] Hugo Carvalho: Yeah. (laughs)

[17:48] Dan Rhatigan: And like so much more investment, so much more effort before I could like start earning money.

[17:53] Hugo Carvalho: Mm-hmm.

[17:53] Dan Rhatigan: And I actually needed to work full-time to pay my bills. So I started working for Type Network, um, and I signed a distribution agreement with them. Um, you know?

[18:08] Hugo Carvalho: Mm-hmm.

[18:08] Dan Rhatigan: So they're, they are just a distributor. They don't own any font IP. Um, they just distribute the fonts and the people who are part of the network. So I s- so Bijou Type really only existed at Type Network at first.

[18:23] Hugo Carvalho: Yeah.

[18:24] Dan Rhatigan: Uh, the way that their setup works is, if you agree that they are your exclusive distributor, you get paid 75% royalties. Um, if you, um, if you distribute elsewhere, you get paid 50% royalties.

[18:42] Hugo Carvalho: Okay. Hmm.

[18:43] Dan Rhatigan: Um, they consider ... If you have your own website, they don't consider that like going against the exclusivity.

[18:50] Hugo Carvalho: Oh, really?

[18:50] Dan Rhatigan: So you can sell things on your own in a Type Network and still get paid the high royalty rate. They also-

[18:55] Hugo Carvalho: That's interesting.

[18:56] Dan Rhatigan: ... would manage, um ... They had an agreement with Adobe to put fonts into Adobe Fonts.

[19:02] Hugo Carvalho: Oh.

[19:03] Dan Rhatigan: Um, not ... I mean, not everything automatically. Like Adobe still has to like throttle the amount of fonts they take on. But they had an agreement with Type Network to take on a certain amount of individual font files every quarter, so you sort of got into the Type Network queue, and that was very helpful. Um, because the Adobe Fonts setup is actually shockingly lucrative.

[19:29] Hugo Carvalho: Yeah, tha- that was one that I was-

[19:30] Dan Rhatigan: Considering h- considering how like-

[19:31] Hugo Carvalho: ... curious about, yeah.

[19:32] Dan Rhatigan: ... small the royalties are, because the scale of use is so big, just the exposure and the money is great.

[19:39] Hugo Carvalho: Yeah, tha- that was one I imagined, uh, uh, that as an alternative model to the more, you know-

[19:45] Dan Rhatigan: Yeah.

[19:45] Hugo Carvalho: ... obvious ones.

[19:46] Dan Rhatigan: Yeah.

[19:46] Hugo Carvalho: Uh, because y- you get paid by syncing, right? So people-

[19:49] Dan Rhatigan: Correct.

[19:49] Hugo Carvalho: ... sync your fonts, you get paid by that. Kind of.

[19:51] Dan Rhatigan: You, yeah, you get paid like cents on the dollar for every 1,000-

[19:57] Hugo Carvalho: Yeah.

[19:57] Dan Rhatigan: ... syncs per month. Um, but it escalates very, very quickly because their user base is so large.

[20:04] Hugo Carvalho: It is, yeah, it is big.

[20:06] Dan Rhatigan: And, and what they consider a sync happens if a user tries a font. Like just while they have something selected, they apply a font. It is not about like what goes into a final project. It really is-

[20:19] Hugo Carvalho: Mm-hmm.

[20:20] Dan Rhatigan: ... has a user ... Did a user activate-

[20:24] Hugo Carvalho: Just-

[20:24] Dan Rhatigan: ... this font for it to appear as text-

[20:25] Hugo Carvalho: Click the button, yeah.

[20:25] Dan Rhatigan: ... in a document. Um, and it stays active for three months, if they don't, if no one ever touches it again. But for any time within a given month that your font is, is active, it goes towards the total. Or if someone has, um, published a web font kit with one of your fonts, you get paid a, a much, much, much, much, much lower rate if it is delivered as a web font. But again, like if there's a f- web te-

[20:57] Hugo Carvalho: It is very cool.

[20:58] Dan Rhatigan: If someone's published a website and Adobe delivers those fonts, every call on that font data counts towards royalties.

[21:09] Hugo Carvalho: Hmmm.

[21:10] Dan Rhatigan: (Coughs) Pardon me.

[21:12] Hugo Carvalho: No, no problem.

[21:12] Dan Rhatigan: Um, the licensing around that is extremely specific. It is broad, but it is very specific. Like, it is a valid license use if someone, um, publishes a website that, uh, where Adobe hosts the fonts. Um-

[21:30] Hugo Carvalho: Oh.

[21:30] Dan Rhatigan: ... if they make something where, um ... if they like, where they made something that is exported out of, you know, one of the Adobe projects.

[21:43] Hugo Carvalho: Like a logo or something?

[21:45] Dan Rhatigan: Yeah, yeah. If you do a logo, if you, you know, you know, make a book, make a poster, make like a little video, anything that sort of like doesn't require distributing the font file itself into the final product.

[22:00] Hugo Carvalho: Mm-hmm.

[22:00] Dan Rhatigan: PDF can be a little bit of a gray area.

[22:04] Hugo Carvalho: S- yeah.

[22:04] Dan Rhatigan: Like in theory it has, you have to like not, you have to like not fully embed the font for it to-

[22:11] Hugo Carvalho: Hmm.

[22:11] Dan Rhatigan: ... be like covered by the licensing. But it just means that I, I ... As you might guess, this comes up a lot with other distributors. Customers say, it is like, "Oh, well, the font was in Adobe, so I'm covered." Well, it is ... You're, the licensing is only covered if you've made the thing entirely within Adobe. Um, you can't like make a website and host the fonts yourself and say that you're covered because Adobe also had the fonts available, or you can't build an app and say like, "Well, I did the mockup, you know, in an Adobe product and then like coded the app." Like the font-

[22:45] Hugo Carvalho: Yeah.

[22:46] Dan Rhatigan: ... file does not travel with the app under Adobe's license.

[22:49] Hugo Carvalho: So, so would you say it doesn't conflict with like getting sales, for example, if you have like ... if you are a foundry and you have your website, for example, and you d- distribute through there and you have also Adobe fonts-

[23:01] Dan Rhatigan: Yeah.

[23:02] Hugo Carvalho: ... and you, you know, distribute through there, do you think it affects sales because people get access to-

[23:06] Dan Rhatigan: I-

[23:07] Hugo Carvalho: ... typefets for free or-

[23:08] Dan Rhatigan: I consider Adobe like my free fo- my free trial service.

[23:13] Hugo Carvalho: Yeah, okay.

[23:13] Dan Rhatigan: You know, to the user, the whole idea of Adobe, of Canva, is it feels free to the user, because the royalty cost is absorbed into the cost of the software that they may be paying.

[23:26] Hugo Carvalho: Yeah.

[23:27] Dan Rhatigan: So, I mean, to the user, like they just have fonts that they can use. Happily, that use is captured and you wind up getting paid. But you get pa-

[23:37] Hugo Carvalho: Yeah, yeah, yeah.

[23:38] Dan Rhatigan: You get paid directly by Adobe for that stuff. Um-

[23:42] Hugo Carvalho: It is actually kind of cool because you get paid by trials almost like-

[23:44] Dan Rhatigan: Yeah.

[23:44] Hugo Carvalho: ... by trials of your fonts.

[23:46] Dan Rhatigan: It is, yeah, exactly. I, I'm very much pro the Adobe Fonts set up for foundries. It has been ... It is also why there's, um ... They have to throttle what they take on so much. There's a cap to what they can pay in royalties. It is directly connected to how much Adobe earns in renewals of Creative Cloud software.

[24:09] Hugo Carvalho: Mm-hmm.

[24:09] Dan Rhatigan: You know, so like the de group, so, you know, the Adobe, the whole Adobe Fonts program, Adobe calculates how much of the Creative Cloud revenue comes down to the value that people get out of Adobe Fonts, and that's what they get as a pool to pay out of. So that's why, you know, they can't just take..... take on everything they want, or everything that people ask to have on there, because they just have a budget of what they can pay.

[24:34] Hugo Carvalho: So, so even if you have an already established foundry, like you, uh, running on Adobe fonts, like, they don't take every single typeface you make?

[24:41] Dan Rhatigan: They do not.

[24:42] Hugo Carvalho: Yeah. Interesting.

[24:44] Dan Rhatigan: Yeah

[24:44] Hugo Carvalho: That's interesting.

[24:45] Dan Rhatigan: And it is, you know, and they have like, you know, their strategy shifts year to year in terms of what they want. So, um, it is great when you can get in there, um, it is frustrating when you can't, uh-

[24:56] Hugo Carvalho: yeah.

[24:56] Dan Rhatigan: ... 'cause it is a nice setup when you're in there.

[24:58] Hugo Carvalho: Yeah.

[24:58] Dan Rhatigan: Um, and then, so, Type Network was my exclusive distributor. Um, they managed the Adobe fonts relationship. I got paid a portion of the royalties that Adobe paid to Type Network, for all the fonts that came in through Type Network. And the w- I mean, the way that was structured was like, you get most of the royalties that Adobe pays, Type Network, just takes a little bit for managing it. Um, I was, um, a little bit concerned at one point that Type Network was going into, um, that they were in sort of a dire financial situation.

[25:39] Dan Rhatigan: Um, so-

[25:39] Hugo Carvalho: Oh.

[25:40] Dan Rhatigan: So, just to, so, actually, to roll back. So, Type Network retail sales through the website, they do offline sales, you know, where people call-

[25:49] Hugo Carvalho: Yeah.

[25:49] Dan Rhatigan: ... them up to negotiate, like, bigger deals. Um, they

take care of all that, you know. Like, they would do marketing for the fonts, not a lot, but they, you know, they do their own marketing. Um, and you're kind of relying on them to do that. And then, all of their, all the fonts that they sell, they do like, a cursory engineering review of, so they also give technical support-

[26:12] Hugo Carvalho: Hmm.

[26:12] Dan Rhatigan: ... when people bring their fonts in for the first time.

[26:13] Hugo Carvalho: That's cool.

[26:14] Dan Rhatigan: Which is all, all-

[26:14] Hugo Carvalho: That's helpful.

[26:15] Dan Rhatigan: All very helpful. Um, the, so, this is gonna be recorded, but I would prefer it if this does not become part of your dissertation bit.

[26:26] Hugo Carvalho: Yeah, yeah, sure.

[26:27] to [30:07] removed per Dan's request.

[30:08] Dan Rhatigan: Um, and so, and it can be, the, the contracts can happen in very, very different ways. Sometimes you really do sell, like, your whole business, the name, the website URL, the social media handles. Sometimes you just sell your fonts and you take the money to, like, reinvest in the next wave of fonts. Like, Sharptime and Font Theory-

[30:28] Hugo Carvalho: Mm-hmm.

[30:30] Dan Rhatigan: ... did not sell-

[30:30] Hugo Carvalho: That's interesting.

[30:31] Dan Rhatigan: ... their businesses. They sold their existing font libraries.

[30:35] Hugo Carvalho: Mm-hmm.

[30:35] Dan Rhatigan: Um, Heffler Linotype, they sold the whole business.

[30:41] Hugo Carvalho: Yeah.

[30:42] Dan Rhatigan: So, so the Type Founders, um, I think they're up to, like, 30 of these libraries now. Um, they've done different kinds of deals. Sometimes it is, like, a slow rollover of complete control, or, you know, sometimes, like, they, they're paying royalties until they pay out the total of the whole agreement, or sometimes it is been cash upfront, but then they take on the designer with an agreement to, like, pay for additional typefaces that get completed.

[31:14] Hugo Carvalho: Yeah.

[31:14] Dan Rhatigan: But the model is transfer the ownership of the font IP.

[31:20] Hugo Carvalho: Hmm.

[31:20] Dan Rhatigan: So slightly different than if you publish through another foundry, very different from you owning your own stuff. Now, it is like

a holding company owns the IP of stuff that's been developed by other people.

[31:33] Hugo Carvalho: Ma- makes sense.

[31:35] Dan Rhatigan: Yeah.

[31:35] Hugo Carvalho: Makes sense.

[31:36] Dan Rhatigan: Uh-

[31:36] Hugo Carvalho: I actually think I talked with, um... Who, who did I talk about that? I think it was Sam Berlow from... Yeah.

[31:45] Dan Rhatigan: Yeah.

[31:45] Hugo Carvalho: Uh, at Fontstand about that (laughs) . I was like-

[31:48] Dan Rhatigan: Yeah.

[31:48] Hugo Carvalho: I was like, who, uh, who told me as well about it? Uh, he was, like, uh, very, very... I only talked with him at, at the bar, and he was like, uh, "Do you want to sell your library?" and I, like, uh-

[31:58] Dan Rhatigan: (laughs)

[31:59] Hugo Carvalho: ... I, I don't have one, but, uh, sure (laughs) .

[32:01] Dan Rhatigan: Yeah. Yeah. It is his job to ask people if they wanna-

[32:04] Hugo Carvalho: Yeah.

[32:04] Dan Rhatigan: ... sell their library.

[32:04] Hugo Carvalho: Of course, of course, of course.

[32:05] Dan Rhatigan: And this is, it is... I think it is very, very funny when people act like Monotype is, like, just going around trying to Hoover up type foundries, uh, knowing that-

[32:20] Hugo Carvalho: Hmm.

[32:20] Dan Rhatigan: ... it is not our primary (laughs) business model. Um, and we turn away people all the time (laughs) who are offering or trying to sell their libraries. We actually, uh-

[32:31] Hugo Carvalho: So, yeah.

[32:32] Dan Rhatigan: ... we actually only do it once in a while.

[32:34] Hugo Carvalho: Online, it is, it is very-

[32:35] Dan Rhatigan: But we just have the resources to do it-

[32:36] Hugo Carvalho: Yeah.

[32:36] Dan Rhatigan: ... in very high-profile situations.

[32:38] Hugo Carvalho: Online, you see, you see a lot of people complaining about Monotype.

[32:42] Dan Rhatigan: Yeah.

[32:43] Hugo Carvalho: Um, I, I think, you know, I, I may-

[32:46] Dan Rhatigan: But look how many, look how many foundries the type founders have bought in the last four or five years.

[32:50] Hugo Carvalho: You know, I, I may be too new to this, and I may be too inexperienced to say this, but I feel like, uh, anyone who helps you make money with something isn't exactly evil. Uh, like, I, I see a lot of people saying, "Oh, Monotype is an evil company." I don't think... They just have a business model. Uh-

[33:08] Dan Rhatigan: Yeah.

[33:09] Hugo Carvalho: ... that's it. If you don't agree, well, you don't have to do business with them, so.

[33:12] Dan Rhatigan: Yeah.

[33:13] Hugo Carvalho: Um...

[33:13] Dan Rhatigan: Well, and that's, and that, that's just it. You know, the... It is the nature, it is the nature of what you agree to when you sign up for distribution or-

[33:24] Hugo Carvalho: Yeah.

[33:25] Dan Rhatigan: ... for someone, someone else to publish your typeface. You are accepting that they have capabilities that you don't to get your font out into the world.

[33:38] Hugo Carvalho: Yeah.

[33:39] Dan Rhatigan: And that's just, you know... And, you know, once a font is made, you know, it is, it is a potential asset. It is something that could generate revenue for you. It is something that could generate revenue for someone else, and there's a variety of different ways that that can happen.

[33:59] Hugo Carvalho: Yeah. Um, but for example, re- regarding the topic of, of marketing-

[34:04] Dan Rhatigan: Yeah.

[34:05] Hugo Carvalho: ... um, of course, you know, i- if you do it through Monotype or whoever, uh, one of those big platforms, you, of course... They take care of that, right?

[34:16] Dan Rhatigan: Mm-hmm.

[34:16] Hugo Carvalho: Of marketing in their own platform's type of thing.

[34:19] Dan Rhatigan: Yeah.

[34:19] Hugo Carvalho: But for example, i- if you... In this case, you have Bisio Type, you have your own foundry. How do you market your typefaces? Of course, everyone knows you, so it is (laughs) kind of like-

[34:29] Dan Rhatigan: Well, I mean, and that's-

[34:30] Hugo Carvalho: ... people already know you, and they kinda-

[34:31] Dan Rhatigan: ... that's-

[34:32] Hugo Carvalho: ... you know?

[34:32] Dan Rhatigan: ... it is kind of what I count on, but the reality is, the people who know me are not the people who are licensing my typefaces.

[34:40] Hugo Carvalho: Oh, that's-

[34:40] Dan Rhatigan: That's-

[34:40] Hugo Carvalho: ... an interesting perspective.

[34:41] Dan Rhatigan: ... that's, I mean, that's, that's the conundrum that I think a lot of type foundries find themselves in, is that, um, the people who know us best are our friends and our peers, they are other type founders. Marketing to, like, a public of, like, other graphic designers or just people from businesses who license fonts is a lot trickier. Um, so-

[35:05] Hugo Carvalho: Yeah.

[35:05] Dan Rhatigan: ... it usually requires quite a lot more resources to do effectively. So when, like, I, I mean, my... So I consider, like, running my website is part of the marketing. Like, I have a place for people to go. I've, like, put effort into, like, the search engine optimization of it, like the way things are tagged in the metadata, how I write about the typefaces-

[35:30] Hugo Carvalho: Yeah.

[35:30] Dan Rhatigan: ... so that search engine crawlers will hopefully, like, bring people-

[35:35] Hugo Carvalho: Yeah.

[35:35] Dan Rhatigan: ... to my website. The more, like, ongoing stuff that I do is, um, social media promotion. I don't pay for it. I just do regular organic stuff by posting about my typefaces and, uh, what I really like to do is find examples of people using my fonts so that I can, like, repost and, um, hopefully-

[35:57] Hugo Carvalho: Yeah, that's cool.

[35:58] Dan Rhatigan: ... encourage them to repost. Um, and that's also why I've opted to continue distribution through a number of platforms. Like, it is... I don't feel like I'm losing money, because distributors are helping people, like, pay for my fonts, so I wouldn't have a way to reach otherwise.

[36:18] Hugo Carvalho: Yeah.

[36:19] Dan Rhatigan: Um, so, you know, like, you know, maybe I don't get, like, the biggest payback if someone has used my fonts through Adobe Fonts, but I know that people are using my fonts there who would never come to my website to license the fonts. They wouldn't have discovered me otherwise. So, like-

[36:36] Hugo Carvalho: Exactly.

[36:37] Dan Rhatigan: ... Chapel Run set designer find- found one of my fonts on Adobe Fonts, I am super excited to let people know that they found that

font and they used it at Coachella. You know, (laughs) it is like-

[36:48] Hugo Carvalho: That's cool, yeah.

[36:50] Dan Rhatigan: Yeah. So like, I, that's why, like, I consider Adobe Fonts, the exposure can just be, like, off the charts.

[36:57] Hugo Carvalho: Off the charts, yeah.

[36:58] Dan Rhatigan: Yeah.

[36:58] Hugo Carvalho: I, I also think, um, op- in this case, o- open source stuff can be-

[37:04] Dan Rhatigan: Yeah.

[37:04] Hugo Carvalho: ... interesting as well. Because-

[37:04] Dan Rhatigan: Yeah.

[37:05] Hugo Carvalho: ... I, I, I went to, um, uh, uh, Google Fonts the other day to look at some numbers of typefaces, and Roboto has like 2.2 billion downloads or something like that.

[37:15] Dan Rhatigan: So-

[37:15] Hugo Carvalho: It is crazy. Uh-

[37:17] Dan Rhatigan: ... you want some crazy examples. So the two typefaces that I've designed that I am least proud of in the world-

[37:24] Hugo Carvalho: (laughs)

[37:25] Dan Rhatigan: ... are on Google Fonts. They're the first things I ever released, whereas like it was at the start of Google Fonts when they were not paying a lot of money to, like-

[37:36] Hugo Carvalho: Hm.

[37:36] Dan Rhatigan: ... have you make something that would be released that way. So I had, like, two weeks to work on each one, and I was not an accomplished enough designer yet to do my-

[37:45] Hugo Carvalho: (laughs)

[37:46] Dan Rhatigan: ... best work in two weeks. But I got them in early. So if you look up Copse, C-O-P-S-E, um, or Astloch, A-S-T-L-O-C-H, um, just look at the stats on these fonts. Like, Copse is one style-

[38:08] Hugo Carvalho: Yeah, it is crazy. (laughs) It is crazy.

[38:11] Dan Rhatigan: ... released in like s-

[38:11] Hugo Carvalho: Like five million, yeah. It is crazy.

[38:13] Dan Rhatigan: Yeah. It is insane the amount of usage that, like, this one style, poorly done typeface gets. But it is just, (clears throat) it was in the right place at the right time to just work its way into projects that keep spinning along.

[38:31] Hugo Carvalho: And-

[38:31] Dan Rhatigan: And I think people, you know, and again, like-

[38:32] Hugo Carvalho: Yeah.

[38:33] Dan Rhatigan: ... it does help a lot and it, that it gets your name circulated. And I think that a lot of fo- a lot of foundries... So marketing is always a challenge, like how much time do you spend on social media? What platforms do you do? Do you pay for ads? What other kind of marketing? Like, Grilly Type does a lot of, like, custom websites-

[38:53] Hugo Carvalho: Microsites, yeah.

[38:54] Dan Rhatigan: ... for specific font families, which are awesome. I don't know what their return on investment is for that. I don't know the people who run Grilly well enough. Um, but, you know, the, um, you know, like I, I printed a specimen poster for my fonts, because I like to print stuff, and I knew that I just wanted to give something to hand out to people-

[39:19] Hugo Carvalho: Yeah.

[39:20] Dan Rhatigan: ... when I see them. Um, the most effective ways in the current media landscape seem to come down to just, like, basics of search engine optimization. That's ultimately what's really gonna drive the needle. Um, you have to have good promo images, you know, so that wherever search engines are scraping information about you from, great to have some images that show off your typefaces, images that include the name of the typeface family and your foundry, um, images that show examples of the fonts in use, even if they're fake.

[40:05] Dan Rhatigan: Um, gen-

[40:05] Hugo Carvalho: Yeah.

[40:05] Dan Rhatigan: ... generally, like, a mix of all of those.

[40:08] Hugo Carvalho: Mm-hmm.

[40:08] Dan Rhatigan: Um, you have to write about your fonts very carefully, you know, like in a concise but compelling way, preferably with some fairly unique language. What is the idea behind the fonts? What is its language coverage? What kind of special-

[40:26] Hugo Carvalho: Mm-hmm.

[40:26] Dan Rhatigan: ... features it have? Does it have, like, distinctive design features worth talking about? Um, and keywords, you know.

[40:36] Hugo Carvalho: Mm-hmm.

[40:36] Dan Rhatigan: So, like, under the hood at Type Network, at My Fonts-

[40:39] Hugo Carvalho: Yeah.

[40:39] Dan Rhatigan: ... in the metadata of my website, there's just keywords about all the fonts, 'cause hopefully you just want search engines to, like-

[40:47] Hugo Carvalho: To pick them up and to help you out.

[40:48] Dan Rhatigan: ... land on your font when someone says like, "I want a f- I want a font that can, that can do this," and that, so that's like... I, I think of that as sort of, like, passive marketing.

[40:59] Hugo Carvalho: Hm.

[41:00] Dan Rhatigan: In a way, it is like you are, you are counting on other machines, other machine systems, to, like, take that and run with it. Um, but it is... it makes a pretty big deal. Um, especially, and that's where, like, you know, being conscious of how your distributors... how s- how well they show up in search results counts for a lot. Like, my fo- if you, like, search for, like, any random font, my fonts is gonna come up. Typetype, the foundry, is clearly putting a lot of money investing in Google searches.

[41:38] Hugo Carvalho: Yeah.

[41:38] Dan Rhatigan: 'Cause even if you search for something that's not from them, a link to their website will come up. Um, and I, I don't have budgets like that. Um-

[41:48] Hugo Carvalho: (laughs)

[41:48] Dan Rhatigan: ... but, you know, but I have-

[41:50] Hugo Carvalho: Yeah.

[41:51] Dan Rhatigan: ... but, like, you know, if you look for Bijou Type, or if you look for me in a search engine, you're gonna get right to me-

[41:58] Hugo Carvalho: Yes.

[41:58] Dan Rhatigan: ... because I've been pretty caref- you know, I, I'm not, like, a marketing powerhouse, but I've just been careful about setting up the families that I have, and I think that's really, really useful. We had, um, we did a business summit for foundries last year at Monotype in Amsterdam, and we were just talking about the importance of this, about how, you know, Monotype doesn't do a lot of direct marketing for specific foundries. Like, we can do campaigns where there'll be in a newsletter, or you can get, like, a banner ad that goes for a while, or if you show up in, like, if you have a new release, it could get into the, you know, hot new fonts.

[42:35] Dan Rhatigan: Um-

[42:35] Hugo Carvalho: Yeah.

[42:36] Dan Rhatigan: ... but what really, really matters in terms of visibility on our platform is..... how are you gonna come up when people search for things? How likely are they to stop when they are just, like, browsing through-

[42:51] Hugo Carvalho: Hmm.

[42:51] Dan Rhatigan: ... um, and finding things. So that's where, like, high impact sample images that communicate something about your family, very, very strong. Like, you can have a set of images. Each can do a slightly different job. Each image oughta be able to do something on its own, 'cause you never know, like, how things may trickle across the internet. But, like, a short, concise, clear description, super useful. Um, you know, and if it is, like if it is not about... And there's different strategies to that. Like, if you, if you don't have the most absolutely unique design, you know, talk about what sets it apart from other geometric sans serifs.

[43:37] Hugo Carvalho: hmmm

[43:37] Dan Rhatigan: Um, or just have, like, a narrative about conceptually how did you arrive at that. Is the narrative about the title of the- the name of the family? Or is the narrative about, like, why you have, um, you know, a family like that? So-

[43:54] Hugo Carvalho: Yeah. So, you know-

[43:55] Dan Rhatigan: Like, it doesn't, it doesn't-

[43:56] Hugo Carvalho: Yeah.

[43:57] Dan Rhatigan: ... fly to just be like, "This is a strong, hardworking sans serif that works at text sizes and has impact at large sizes. It is good for magazines or stationary." That's like the most, uh, totally accurate perhaps, but the most generic way to describe a family. And if it is generic, it is not gonna help people find you, unfortunately.

[44:19] Hugo Carvalho: Yeah. Um, do you think that, uh, personalization helps maybe?

[44:24] Dan Rhatigan: I-

[44:25] Hugo Carvalho: Be- because of-

[44:26] Dan Rhatigan: ... I totally do.

[44:27] Hugo Carvalho: Uh, I'm, I'm going to give an example-

[44:29] Dan Rhatigan: Yeah.

[44:29] Hugo Carvalho: ... a very practical example, because, for example, I'm creating a typeface which is called Avelino.

[44:34] Dan Rhatigan: Okay.

[44:34] Hugo Carvalho: And, and this is because you, you talked about the name.

[44:37] Dan Rhatigan: Yeah.

[44:37] Hugo Carvalho: Uh, and I'll, I'll... I'm working on my first typeface, which is going to be released, I don't know when, whenever I finish-

[44:42] Dan Rhatigan: Yeah.

[44:42] Hugo Carvalho: ... my master degree and I have time to finish it up.

[44:44] Dan Rhatigan: Yeah.

[44:45] Hugo Carvalho: Uh, and it is called Avelino because it is named after my grandfather.

[44:48] Dan Rhatigan: Yeah.

[44:49] Hugo Carvalho: And it, it has kind of a backstory with that. And because it has that, you, you believe by sticking to the basics of marketing and stuff like that, you can pull through that personality and eventually people will

connect to that.

[45:05] Dan Rhatigan: Yeah.

[45:05] Hugo Carvalho: Okay.

[45:05] Dan Rhatigan: Yeah, exactly. And also-

[45:07] Hugo Carvalho: Yeah.

[45:07] Dan Rhatigan: ... and, you know, and I think f- once you have a few typefaces available, thinking about the narrative around your whole library is also just as important.

[45:20] Hugo Carvalho: Hmm.

[45:20] Dan Rhatigan: You know, like what is, what is the narrative of your foundry? What is the s- what is the story of what you do and why you make typefaces? Um, and, and-

[45:30] Hugo Carvalho: Yeah. I, I see that through a lot of foundries, of course.

[45:34] Dan Rhatigan: Yeah.

[45:34] Hugo Carvalho: And what, uh, wh- why do you make fonts, but also what type of fonts do you make kind of?

[45:38] Dan Rhatigan: Yeah. Exactly.

[45:39] Hugo Carvalho: Um...

[45:39] Dan Rhatigan: And, and in reality, so I've... I have talked to a lot of type foundries over the years. I've noticed some patterns, and I've started really, like, explicitly asking them about this. And everyone kind of confirms my theory.

[45:55] Hugo Carvalho: Hmm.

[45:55] Dan Rhatigan: Every foundry, at some point, gets to the point where they have, like, one family that is carrying most of the weight.

[46:06] Hugo Carvalho: Hmm.

[46:06] Dan Rhatigan: Sometimes, like, they're lucky and it is like two or three once you're really established. But there's usually, like, one family that is generating most of the revenue. It is almost always a sans-serif.

[46:18] Hugo Carvalho: (laughs)

[46:19] Dan Rhatigan: Um, you know, a multi-weight sans, 'cause in, in the end, that is what most people need for-

[46:27] Hugo Carvalho: Hmm.

[46:27] Dan Rhatigan: ... work that people do at scale. Um-

[46:31] Hugo Carvalho: Yeah.

[46:31] Dan Rhatigan: ... but very often, it is the interesting stuff that makes you stand out of the pack. So, like, your crazy display faces, your high concept

narratives or families, the things that may not be practical, but are different, are really truly different from what other foundries have.

[46:54] Hugo Carvalho: Hmm.

[46:54] Dan Rhatigan: Um, and once people, you know, are drawn to you and your library because of that, you know, maybe the credit card comes out because they see like, you know, it is like, "I want this sans. I like this personality of this whole project, so let's recommend this." Because really, when designers have to pitch work that they make, they often have to lean into the narrative you've set up-

[47:19] Hugo Carvalho: Yeah.

[47:19] Dan Rhatigan: ... to support the typefaces they're recommending someone pay for to go into a project. So, I think this idea of having a narrative around your typefaces, around the rest of your library, about the foundry is also a way of, like, supplying someone else to endorse those for you. You're, like, feeding something that can make-

[47:43] Hugo Carvalho: That-

[47:43] Dan Rhatigan: ... the whole project stay-

[47:44] Hugo Carvalho: ... those qualities, yeah.

[47:46] Dan Rhatigan: Yeah. So, and it is, it is, like, kind of unsexy to think about, but, like, good writing helps sell good typefaces, because it-

[47:55] Hugo Carvalho: It makes sense.

[47:56] Dan Rhatigan: It is part of, like, the whole thing that you're setting up for someone else to, like, select your typeface and get someone else on board with using it.

[48:07] Hugo Carvalho: Hmm.

[48:07] Dan Rhatigan: Like the Bijou projects, I haven't, I have not finished my Workhorse Sans Serif yet, and I'm really, really eager to, because I wanna see if this experiment holds true to me. But I started releasing-

[48:21] Hugo Carvalho: It, it feels good to be like your best release. (laughs) Yeah.

[48:23] Dan Rhatigan: Yeah. Like, I started releasing display typefaces because I wanted to get out the stuff that would make a splash, you know?

[48:30] Hugo Carvalho: Mm-hmm.

[48:30] Dan Rhatigan: The stuff I could just, like, wrap it up, clear it, move it out, start generating interest. And yeah, it is like I'm... You know, it is like they're starting to bring in money. And I can see from the people who are using the typefaces that they are connecting to the narrative of the typefaces-

[48:48] Hugo Carvalho: Hmm.

[48:48] Dan Rhatigan: ... which is great. It is like, it is... That's very, very much what I want, you know? The, the projects that I vers- leased first, I sat on because I could have had Monotype release them, I could have had Adobe release them, but..... they were not narratives that belonged in either of those

companies. They made sense once Bijou was this small little boutique foundry available to catch them.

[49:12] Hugo Carvalho: Mm-hmm. Yeah.

[49:13] Dan Rhatigan: And like, I love being surprised by where my fonts show up. Um, but I really love when I can tell that people have chosen the fonts because the story of the fonts relates somehow to the story of the people using the fonts. Which means that's... I wanna have a text face that they can use for like, you know, some other thing that's connected to it.

[49:37] Hugo Carvalho: Yeah. That's interesting. That's very interesting. Um, h- how do you go about adding typeface to your catalog? Like, new releases and stuff like that? Uh, I imagine a lot of them is true to your own, you know, ideas and concepts that are in your head for so long?

[49:56] Dan Rhatigan: Yeah. I mean, that's the thing. It is like the, you know, custom type fo- projects go really fast 'cause you have to just, like... you've got a brief, you've gotta commit to a decision, you've gotta make it happen.

[50:08] Hugo Carvalho: Yeah.

[50:08] Dan Rhatigan: Um, retail stuff, for me at least, really has to simmer for a while. Um, the typefaces... 'cause even if there's like, a strong conceptual idea, I don't work full time on designing typefaces. So like, I do them in bits and pieces so the idea-

[50:27] Hugo Carvalho: Yeah.

[50:28] Dan Rhatigan: ... kind of tends to evolve a little bit. Like, the... I mean, the f- the first things, the display typefaces, it was... the time was just a matter of like, finishing them as the, you know, the scale ballooned, but the idea was very clear. I have a slab serif family, um, that, that took a longer time to come together because it is like I had to find a wandering path through the wood. 'Cause what I really wanted was...

[50:59] Dan Rhatigan: I knew the qualities it needed to have-

[51:02] Hugo Carvalho: Hm. Yeah.

[51:04] Dan Rhatigan: ... but it took a while to break that down into like, very specific shape decisions, and then to get it to, like, the level of refinement that I needed, you know? I knew I wanted it to be, like... I knew I wanted a slab serif that was good for reading text because it was... would be appropriate for magazines and... which is conceptually part of the Bijou story, is like, magazine publishing.

[51:27] Hugo Carvalho: Yeah.

[51:27] Dan Rhatigan: Um, but, you know, it is like I was beginning with... I want it to feel a little warm, I don't want it to feel really geometric like Rockwell. But that negotiation took a lot of time to, like, land. Um, and now I'm in the process of trying to wrap up the italic so that people will buy the family.

[51:48] Hugo Carvalho: (laughs)

[51:49] Dan Rhatigan: Um, because if you have a family that works for text, people really, really want italics that go with it. Um, and so I'm following... I'm trying to l- it is like I'm trying to find a parallel path conceptually that makes sense for the italic that goes with this other family that I designed.

[52:06] Hugo Carvalho: Mm-hmm.

[52:07] Dan Rhatigan: And meanwhile, I am also trying to, like, finish a sans-serif, which has to go through the same process of like, "Well, why do I need to have... I know why I need a sans-serif, but like, why does the world need another sans-serif, and why should it be coming from me?" is another nebulous design thing that I have to think through.

[52:28] Hugo Carvalho: Yeah. why do you want to do it?

[52:31] Dan Rhatigan: But... and it takes... and, and also for me, because the, the projects have to happen slowly 'cause it is very part-time, um, it has to be an idea that remains interesting enough to me to justify that I sit with it for a few years.

[52:47] Hugo Carvalho: Mm-hmm.

[52:47] Dan Rhatigan: So like, you know, a sans-serif can be so boring to work on if you're not excited about it. So, I have to really keep zeroing in on what's an interesting task.

[52:59] Hugo Carvalho: Yeah. M- that makes a lot of sense (laughs) . Um-

[53:02] Dan Rhatigan: And then... and a- and along the way, that's also when the narrative develops. Because I'm constantly trying to justify, you know-

[53:10] Hugo Carvalho: Why am I doing this? Why is this cool?

[53:10] Dan Rhatigan: ... the effort that I'm putting into it. And, you know, those justifications become part of the story. Like, "This is why I believe this needs to exist in the world."

[53:20] Hugo Carvalho: Hm. Yeah, it makes a lot of sense.

[53:23] Dan Rhatigan: Yeah.

[53:24] Hugo Carvalho: Um, well, um, yeah, uh, I mean, pretty much w- all the questions I think I had are answered. I don't know if you wanna say anything else, uh, about it, but I think like, mo- the core questions that I had-

[53:38] Dan Rhatigan: I think I've hit most of them.

[53:39] Hugo Carvalho: Eh?

[53:39] Dan Rhatigan: I think I've hit most of what I've-

[53:40] Hugo Carvalho: Yeah. I, I think like-

[53:40] Dan Rhatigan: ... tried to distill down to cover the....

[53:43] Hugo Carvalho: All, all of the core ideas are said. So, yeah (laughs).

[53:48] Dan Rhatigan: (laughs)

[53:48] Hugo Carvalho: Um, thank you. You know, uh, it is-

[53:51] Dan Rhatigan: Thanks. All right.

[53:51] Hugo Carvalho: It is always very interesting.

[53:52] Dan Rhatigan: My pleasure. That'll be, that'll be €5,000 (laughs) .

[53:55] Hugo Carvalho: Oh, I... you know, honestly, I wish I had money to pay you.

[53:59] Dan Rhatigan: (laughs)

[53:59] Hugo Carvalho: If I had it, I wouldn't mind to pay you.

[54:01] Dan Rhatigan: (laughs)

[54:01] Hugo Carvalho: Because it is truly like, um... you know, as someone who's starting out, it is very interesting to find mentors, uh, that in many ways, they are not official mentors. You are not me-

[54:14] Dan Rhatigan: Yeah.

[54:15] Hugo Carvalho: ... mentoring me officially or something like that.

[54:17] Dan Rhatigan: Yeah.

[54:17] Hugo Carvalho: But, uh, it, it is interesting to find mentors throughout the development of a career and stuff like that. So, I think it is very cool.

[54:24] Dan Rhatigan: Yeah. Well, I'm-

[54:24] Hugo Carvalho: Um, and it is very meaningful to me. So, I think it is a-

[54:27] Dan Rhatigan: I mean, I'm, I'm conscious of the fact that I've had a lot of good breaks professionally. I've, I picked up a lot. I, I mean, I, I think there's a handful of people who have seen the amount of things that I have.

[54:44] Hugo Carvalho: Mm-hmm.

[54:44] Dan Rhatigan: I've had the good fortune of working in a lot of different places. Um, it would be wrong for me not to share this (laughs) when I have the opportunity. You know, it is like the...

[54:56] Hugo Carvalho: Yeah. Uh-

[54:57] Dan Rhatigan: I have a, I have a-

[54:58] Hugo Carvalho: I mean, yeah.

[54:59] Dan Rhatigan: ... I have a, I have a lot of perspective that can help, and I like to share it as much as possible as much as

[55:04] Hugo Carvalho: Uh-It also comes down to, like, um, when I went to, to Fontstand, my approach... Although some people interpreted it as marketing (laughs), which I thought was funny. But my approach was more from the, "Hey, here am I. I'm fairly shy. Nice to meet you. And what can I... can you teach me in five minutes," kind of.

[55:23] Dan Rhatigan: Yeah. (laughs)

[55:23] Hugo Carvalho: And I, I ended up being kind of a sponge, which was exactly what I wanted to do. I want to be a sponge for... of information.

[55:30] Dan Rhatigan: Yeah.

[55:30] Hugo Carvalho: Um, and that worked fairly well. I don't know exactly... Um, ah, I don't know exactly how this is going to turn out in terms of like a career or something like that.

[55:39] Dan Rhatigan: Yeah.

[55:39] Hugo Carvalho: Like, I am not... You know, I don't know my path forward. I'm kind of figuring that out as I go. (laughs) Uh, but I think it is very cool, like, to, to meet people, to talk to people, to interact-

[55:51] Dan Rhatigan: Yeah.

[55:51] Hugo Carvalho: ... to put myself out there, which I'm a bit... always afraid of doing. But it is, it is-

[55:56] Dan Rhatigan: Yeah.

[55:56] Hugo Carvalho: It has been a cool process.

[55:58] Dan Rhatigan: Yeah.

[55:58] Hugo Carvalho: Um...

[55:59] Dan Rhatigan: Well, and you know, even if you don't release your own typefaces, you know, if you go into like a field of design where you're just choosing typefaces, I think you've learned a lot to chew on in terms of like, whose work do you wanna support? Like, where-

[56:17] Hugo Carvalho: Yeah.

[56:17] Dan Rhatigan: ... where does it make sense for you to, like, get your licensing? Do you wanna like... What licensing models, you know, are aligned with your, your business goals? Um...

[56:28] Hugo Carvalho: Yeah. E-

[56:29] Dan Rhatigan: It is like, it is, it is useful-

[56:31] Hugo Carvalho: Hummm

[56:31] Dan Rhatigan: You know, the, the idea of sharing what we know is not just about like raising up other type designers. It can also be very much about like raising future customers and future clients and...

[56:49] Hugo Carvalho: Yeah.

[56:49] Dan Rhatigan: And, and making a living I guess.

[56:49] Hugo Carvalho: Yeah, I definitely see, see that, uh... see what you mean with that. Um, I, I also think, um, it is a beautiful way of thinking about it. Um, you know, I teach (laughs) so-

[57:04] Dan Rhatigan: Yeah.

[57:04] Hugo Carvalho: Uh, b- beyond just, uh, designing, I also teach. Uh, although it is like secondary level education.

[57:10] Dan Rhatigan: Yeah.

[57:10] Hugo Carvalho: Um, and it is very cool to pass on knowledge. Although I never believe, and I'll probably never believe that I hold all the knowledge in a specific field. I'm sure you feel that way as well, uh, that you have so much to learn.

[57:24] Dan Rhatigan: Yeah.

[57:25] Hugo Carvalho: Uh, but it is very cool when you can like share stuff with people that wanna listen to you (laughs), which uh-

[57:32] Dan Rhatigan: Yeah.

[57:32] Hugo Carvalho: ... in my case, I'll, I'll admit it is like one in 20 students, but uh-

[57:37] Dan Rhatigan: (laughs)

[57:37] Hugo Carvalho: (laughs) So I have that difficulty going on for me, but-

[57:41] Dan Rhatigan: I think that's, I think that's what most teachers discover. (laughs)

[57:44] Hugo Carvalho: Discover, right? Uh, but I, I think it is cool to, to you know, learn this and try to find mentors and people that can teach you. Um, and it is very cool that you did this for me, and, uh, I'm very happy. So... (laughs) Yeah.

[58:00] Dan Rhatigan: It was, it was totally a pleasure. Um, and you know, if-

[58:03] Hugo Carvalho: And uh, you know, I, I wish I could learn more from you-

[58:04] Dan Rhatigan: As you like write this-

[58:05] Hugo Carvalho: ... because it is insane-

[58:07] Dan Rhatigan: Yeah.

[58:07] Hugo Carvalho: ... the information that you just shared.

[58:08] Dan Rhatigan: Yeah. Well, like... And as, as you write stuff up, you know, like writing is very much part of the process of getting to the end. Um, feel free to reach out if you have like follow-up questions or stuff, you know. Don't be shy.

[58:20] Hugo Carvalho: Yeah. Yeah. Awesome. Right now, I can give you like a few updates on, on my dissertation actually-

[58:27] Dan Rhatigan: Yeah.

[58:27] Hugo Carvalho: ... if you wanna listen to them.

[58:27] Dan Rhatigan: Yeah.

[58:28] Hugo Carvalho: Uh, so my, my initial, um, proposal was to... I... Essentially these researchers from the University of Johns Hopkins, they created a tool called VOCR.

[58:39] Dan Rhatigan: Mm-hmm.

[58:40] Hugo Carvalho: And, uh, this tool is essentially a pipeline. And, and what that pipeline allows you to do, in theory, is to simulate human vision for you to test your typefaces-

[58:51] Dan Rhatigan: Mm-hmm.

[58:52] Hugo Carvalho: ... um, legibility, in terms of legibility and stuff like that.

[58:56] Dan Rhatigan: Yeah.

[58:56] Hugo Carvalho: So that's like the goal, which would be cool because, for example, people like Octavio who developed like the typeface, Tiphares, I think that's the proper name, um, they... You know, he did not... Ended up testing the typeface, and it would be a, a cool way for people to test their typeface, although-

[59:15] Dan Rhatigan: Yeah.

[59:15] Hugo Carvalho: ... human testing is always needed. So there's that.

[59:18] Dan Rhatigan: Yeah.

[59:18] Hugo Carvalho: And what I've done, I have done... Essentially picked up their tool. I'm trying to understand it fully, I'm trying to understand how to properly like implement a tool and how to... Analyzing problems that I had, which I had, uh, uh, a bunch of them-

[59:34] Dan Rhatigan: Mm-hmm.

[59:34] Hugo Carvalho: ... uh, trying to, you know, run their own scripts and stuff like that. And then it, it is about... It was about running a test protocol based on the Kevin Larsson's methodologies for testing-

[59:47] Dan Rhatigan: Yeah.

[59:47] Hugo Carvalho: ... with Sitka, the Sitka typeface-

[59:49] Dan Rhatigan: Mm-hmm.

[59:49] Hugo Carvalho: ... uh, which used the staircase procedure. Um, and adapting that and trying to test out with visually impaired people and seeing if the tool and the visually impaired people will produce the same type of legibility metrics under cer- like five tester typefaces.

[01:00:05] Dan Rhatigan: Yeah.

[01:00:05] Hugo Carvalho: And yeah, like this has been it in terms of like what my dissertation is. It is very technical. I, I was a bit... I, I think I bit, uh, a little more than I could chew (laughs) -

[01:00:16] Dan Rhatigan: (laughs)

[01:00:17] Hugo Carvalho: ... I'm going to be very honest with you, because, um, i- if, if I could slap m- m- myself in the past, I would probably do that-

[01:00:24] Dan Rhatigan: (laughs)

[01:00:24] Hugo Carvalho: ... and I would say, "Dude, just make, just make a typeface." It is so much like-

[01:00:30] Dan Rhatigan: Oh, like it is so easy? (laughs)

[01:00:32] Hugo Carvalho: No, no. It, it is not like it is so easy, like it takes time. Like I... I don't know if it is just me, but I'm like really slow at, uh, de-signing typefaces apparently. I've been working on this one for like a year or something, and it is like throwing me off so much (laughs) 'cause you are like, "Oh, you draw a glyph," and you are... and then you are like, "Oh man, I have to go back (laughs) -

[01:00:54] Dan Rhatigan: (laughs)

[01:00:54] Hugo Carvalho: ... and redo stuff." And then you find problems. Um, and it is, it is cool. I, I think it is a very iterative process and very interesting process. Uh, but yeah, you end up, you know, finding yourself a little bit lost. Uh-

[01:01:06] Dan Rhatigan: Yeah.

[01:01:07] Hugo Carvalho: But it is cool. It is cool. And yeah, um, right now I am essentially doing the final push to have everything ready for the end of May... so that at the end of May, I can, uh, deliver the full dissertation, um, as it is to my supervisor. He can comment and then I can do the pagination for it and send it off. So-

[01:01:30] Dan Rhatigan: All right.

[01:01:31] Hugo Carvalho: ... yeah.

[01:01:31] Dan Rhatigan: Cool.

[01:01:32] Hugo Carvalho: Crunching the data as we speak, so the data that I got from the visually impaired participants and the tool itself, I'm kind of comparing it. There seems to be a correlation within a certain margin of error, which means that in theory, if they improve their tool, it can be used for the purpose that they intend, which is to simulate-

[01:01:52] Dan Rhatigan: Yeah.

[01:01:53] Hugo Carvalho: ... visual impaired vision. Uh-

[01:01:55] Dan Rhatigan: Awesome.

[01:01:55] Hugo Carvalho: ... it was also very interesting to talk to... and if you have thoughts on this, please add... to talk to people like Erik van Blokland and, um, Erik Spiekermann and stuff like that, about, for example, how this tool... and Dave Crossland and others, how this tool can be implemented on the, um, on the market.

[01:02:11] Dan Rhatigan: Yeah.

[01:02:12] Hugo Carvalho: Not in terms of like commercially speaking, but in terms of more like in your own methodologies. A lot of people, I think, don't do a lot of legibility testing from what I discerned from my own, um, questions that I posed to people. So a lot of people don't do a lot of legibility testing. They don't really focus on that aspect.

[01:02:34] Dan Rhatigan: Yeah.

[01:02:34] Hugo Carvalho: But I think it would be cool if like, for example,

like the... I don't know if you use Glyphs or not, but, uh, in Glyphs you have that slider that you can simulate blur.

[01:02:42] Dan Rhatigan: Yeah.

[01:02:42] Hugo Carvalho: It would be kind of cool if you ha- if you had a version of that that would simulate like different types of visual impairments.

[01:02:47] Dan Rhatigan: Yeah. I mean, it would... I think everyone... I think people think about legibility when they design often, but-

[01:02:58] Hugo Carvalho: Mm-hmm.

[01:02:58] Dan Rhatigan: ... um, you know, they are digesting commonly understood principles, but testing is slow and expensive, so it is not-

[01:03:06] Hugo Carvalho: Yeah.

[01:03:06] Dan Rhatigan: ... it is hard to like build it into regular design process. A tool would be like hugely popular if it were accessible. Yeah.

[01:03:14] Hugo Carvalho: Yeah. That, that was my exact thought. (laughs) Tha- that's like wh- why I'm interested in this. Like I was just, um... Before I started writing this and going into this idea, I, I was like, "Man, I'm interested in legibility. I'm new- I'm interested in visual impairments, um, also because I have high degree of myopia." And I'm like, "It would be cool if a tool existed that allowed you to test with people." And I ended up finding this, uh, uh, research paper in... that was released in 2025, which means this is like incredibly fresh. And I've been really lucky because not only, you know, from being very integrated in the type design community, i- it has been cool, but also the fact that the re- the researchers, the original researchers came through and like, uh, talked with me and stuff like that. That's always interesting because you add up to an international conversation that is happening.

[01:04:06] Hugo Carvalho: And so, yeah, this dissertation more than being my thing, it is a global thing and it can contribute to something-

[01:04:14] Dan Rhatigan: Yeah.

[01:04:15] Hugo Carvalho: ... hopefully interesting.

[01:04:16] Dan Rhatigan: Well, and, you know, part, part of the reason that, you know, workhorse sans-serifs are very popular and sell well, um, is they, you know... Legibility is a really, really big factor in people like investing in typefaces. You know-

[01:04:37] Hugo Carvalho: Yeah.

[01:04:37] Dan Rhatigan: ... for the people who are like laying down money for proper licensing, legibility is one of those things that like, they kind of need to stand by and say like, "Yes, this will be effective. This will help us communicate." So, you know, a tool that does testing to help like really, you know, vouch for that quality in a typeface, it is like a design aid, but it is also part of the narrative, it is part of a sales tool. It would be very valuable to people. Like if it just like... you know, if it can like produce any kind of verifiable data, that's, that's gold.

[01:05:13] Hugo Carvalho: Yeah. Well then, I pretty much don't have anything else to, to... I, I could talk to you for hours, you know this-

[01:05:20] Dan Rhatigan: (laughs)

[01:05:20] Hugo Carvalho: ... but I don't want to steal more of your time. I think you have been very kind with me and I don't wanna-

[01:05:24] Dan Rhatigan: Yeah.

[01:05:24] Hugo Carvalho: ... you know-

[01:05:24] Dan Rhatigan: No, no.

[01:05:25] Hugo Carvalho: ... be abusive of that.

[01:05:26] Dan Rhatigan: It was fun. No, I gotta go get some dinner. But it was very fun to talk to you. I'm glad, I'm glad I was able to help.

[01:05:31] Hugo Carvalho: Hey-

[01:05:31] Dan Rhatigan: Um-

[01:05:31] Hugo Carvalho: ... it was awesome.

[01:05:32] Dan Rhatigan: All right.

[01:05:32] Hugo Carvalho: Thank you so much.

[01:05:33] Dan Rhatigan: I'll, uh, I'll catch you again sometime, man. Good luck with all the work.

[01:05:35] Hugo Carvalho: I'll catch you again sometime as well.

[01:05:37] Dan Rhatigan: Yeah. Cool.

[01:05:37] Hugo Carvalho: It was a pleasure.

[01:05:38] Dan Rhatigan: All right. Take care.

[01:05:39] Hugo Carvalho: Take care.

This dissertation investigates the use of vision impairment simulation technologies for evaluating typeface legibility. By comparing results from the VI-OCR simulation pipeline with tests involving participants with visual impairments, the study demonstrates that computational models can reliably replicate human reading performance. The findings support the use of simulation-based methodologies as efficient tools for developing and assessing accessible typefaces.

Author: Hugo Carvalho

Master's Supervisor: Professor Pedro Manuel Reis Amado, Ph.D.

Dissertation submitted in partial fulfillment of the requirements for the Master's Degree in Graphic Design and Editorial Projects.

FBAUP, June, 2026