U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

# INTELIGÊNCIA ARTIFICIAL AO SERVIÇO DO ARQUIVO, DA IDENTIDADE E DA MEMÓRIA

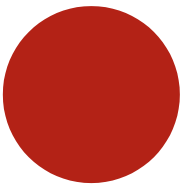
**Sara Filipa Gonçalves Esteves**

M

**2026**

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**
**FACULDADE DE LETRAS DA UNIVERSIDADE DO PORTO**

# Inteligência Artifficial ao Serviço do Arquivo, da Identidade e da Memória: a importância dos metadados na recuperação da informação em retratos de estúdio

**Sara Filipa Gonçalves Esteves**




Mestrado em Ciência da Informação

Orientador: Professora Doutora Joana Patrícia de Sousa Rodrigues

Junho de 2026

© Sara Filipa Gonçalves Esteves, 2026

# Inteligência Artifficial ao Serviço do Arquivo, da Identidade e da Memória: a importância dos metadados na recuperação da informação em retratos de estúdio

**Sara Filipa Gonçalves Esteves**

Mestrado em Ciência da Informação

Aprovado em provas públicas pelo Júri:

Presidente: Professora Doutora Carla Alexandra Teixeira Lopes

Arguente: Professora Doutora Maria Cristina Vieira de Freitas

Orientador: Professora Doutora Joana Patrícia de Sousa Rodrigues

---

Julho 10, 2026

# Resumo

A presente investigação observa o contributo da Inteligência Artificial na descrição e recuperação de fotografias, a partir de retratos de estúdio de uma coleção do Arquivo Municipal de Viana do Castelo, em que a problemática assenta na necessidade de produzir descrições mais consistentes, normalizadas e orientadas para facilitar a recuperação da informação.

O objetivo consistiu em comparar a descrição automática do Módulo de Descrição de Imagem, a descrição humana/manual e um teste complementar com o ChatGPT Vision. Para se conseguir atingir o objetivo, desenvolveu-se uma metodologia que incluiu análise arquivística, criação de lista de termos controlados, etiquetagem de imagens, comparação de resultados e validação humana.

No final, os resultados mostram que o Módulo de Descrição de Imagem produziu etiquetas genéricas, enquanto a descrição humana foi mais exaustiva e precisa, aumentando a média de etiquetas por imagem, de cerca de 4 para 22. O ChatGPT Vision apresentou melhores resultados quando orientado por instruções e termos controlados, ajustando a descrição aos critérios fornecidos.

Verificou-se que a Inteligência Artificial pode apoiar a descrição e a recuperação da informação em fotografias, desde que seja integrada num processo supervisionado, com intervenção humana, sobretudo quando estão em causa a interpretação do contexto, o uso consistente dos termos, a validação dos elementos identificados e a produção de descrições adequadas à recuperação da informação.

**Palavras-chave:** Descrição; Inteligência Artificial; Fotografia de estúdio; Recuperação da Informação; Vocabulário Controlado.

# Abstract

This research examines the contribution of Artificial Intelligence to the description and retrieval of photographs, using studio portraits from an AMVCT collection, where the challenge lies in the need to produce more consistent, standardised, and targeted descriptions to facilitate the retrieval of information.

The aim was to compare the automatic descriptions generated by the Image Description Module (MDI) with human/manual descriptions, and to conduct a supplementary test using ChatGPT Vision. To achieve this aim, a methodology was developed that included archival analysis, the creation of a list of controlled terms, image tagging, comparison of results, and human validation.

Ultimately, the results show that the Image Description Module produced generic tags, whilst the human description was more comprehensive and accurate, increasing the average number of tags per image from around 4 to 22. ChatGPT Vision performed better when guided by instructions and controlled terms, tailoring the description to the criteria provided. In conclusion, Artificial Intelligence can support the description and retrieval of information in photographs, provided it is integrated into a supervised process involving human intervention, particularly when it comes to interpreting context, the consistent use of terms, the validation of identified elements, and the production of descriptions suitable for information retrieval.

**Keywords:** Artificial Intelligence; Controlled Vocabulary; Description; Information Retrieval; Studio Photography.

# NU Objetivos de Desenvolvimento Sustentável

Os Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas constituem um quadro global de referência para a construção de um futuro mais sustentável. A presente dissertação contribui para esse enquadramento ao analisar a inovação digital aplicada à descrição arquivística, a salvaguarda do património fotográfico e a melhoria do acesso à informação pública.[1]

Os Objetivos de Desenvolvimento Sustentável considerados no âmbito desta dissertação são os seguintes:

**ODS 9** Indústria, inovação e infraestruturas: construir insfraestruturas resilientes, promover uma isndustrialização inclusiva e sustentável, e fomentar a inovação.

**ODS 11** Cidades e comunidades sustentáveis: tornar as cidades e as comunidades inclusivas, seguras, resilientes e sustentáveis.

**ODS 16** Paz, justiça e instituições sustentáveis: promover sociedades pacíficas e inclusivas, proporcionar o acesso à justiça para construir instituições eficazes, responsáveis e inclusivas.

| **ODS** | **Meta** | **Contributo** | **Indicadores de desempenho e métricas** |
|---|---|---|---|
| 9 | 9.5 | Avaliação de ferramentas de Inteligência Artificial enquanto apoio à descrição arquivística de imagens e ao enriquecimento dos metadados. | Comparação entre descrições automáticas, assistidas e humanas; número, especificidade e granularidade dos pontos de acesso. |
| 11 | 11.4 | Descrição e divulgação da Coleção Foto Roriz, contribuindo para a salvaguarda, valorização e acessibilidade do património fotográfico local. | Construção de vocabulário controlado, elaboração de um guia de descrição e aplicação de metadados normalizados à amostra selecionada. |
| 16 | 16.10 | Melhoria da recuperação da informação e do acesso público às fotografias do Arquivo Municipal, através de metadados consistentes, normalizados e validados. | Precisão dos resultados recuperados e refinamento progressivo das pesquisas através da combinação de termos controlados. |

[1] Nações Unidas, Objetivos de Desenvolvimento Sustentável: https://sdgs.un.org/goals

# Agradecimentos

Há momentos em que olhamos para trás e mal reconhecemos o caminho que percorremos. Chegar até aqui custou noites, custou dúvidas, custou escolhas difíceis e recomeços que ninguém viu. Por isso, estes agradecimentos são um reconhecimento genuíno, sentido e profundo de que este caminho foi construído com muito mais do que palavras minhas.

À minha orientadora, agradeço a disponibilidade, a paciência e a orientação nos momentos em que o caminho parecia menos claro, as suas palavras e o seu acompanhamento fizeram toda a diferença.

Aos professores que me desafiaram a pensar mais longe, a questionar e a melhorar, deixo também o meu agradecimento.

Aos colegas que partilharam comigo dúvidas, cansaço e pequenas conquistas, obrigada pela companhia, pela escuta e pelo incentivo.

Ao Município de Viana do Castelo, à Mind e a todos os profissionais que contribuíram para este estudo, agradeço a disponibilidade, a colaboração e a partilha de conhecimento, sem as quais este trabalho não teria sido possível.

À minha família, que esteve presente nos dias bons e, sobretudo, nos dias difíceis, o meu agradecimento mais profundo. Obrigada por me apoiarem quando duvidei, por aceitarem os silêncios, o cansaço e a ausência, e por nunca deixarem de acreditar em mim.

A todos os que, de alguma forma, deixaram uma marca nesta caminhada, o meu sincero Obrigado!

Sara Esteves

*"… depois logo se vê!"*

Helena Neves

# Sumário


**1 Introdução** **1**

1.1 Contextualização . . . 1
1.2 Motivação . . . 2
1.3 Enquadramento do projeto . . . 3
1.4 Objetivos . . . 3
1.5 Problemas identificados e a resolver . . . 4
1.6 Estrutura da Dissertação . . . 5

**2 Revisão Sistemática de Literatura** **6**

2.1 Enquadramento conceptual . . . 6
2.1.1 Fotografia . . . 6
2.1.2 Arquivos fotográficos . . . 7
2.1.3 Metadados . . . 7
2.1.4 Esquemas de metadados . . . 8
2.1.5 Indexação . . . 9
2.1.6 Recuperação da informação . . . 9
2.2 Metadados em fotografias . . . 10
2.2.1 Funções dos metadados em fotografias . . . 10
2.2.2 Metadados descritivos . . . 11
2.2.3 Esquemas e normas aplicadas às fotografias . . . 12
2.2.4 Limitações da descrição tradicional . . . 12
2.3 Inteligência Artificial: automatização de metadados . . . 13
2.3.1 Inteligência Artificial nos Arquivos . . . 13
2.3.2 Classificação de imagem e automatização da extração de metadados . . . 14
2.3.3 Integração de IA com esquemas de metadados . . . 15
2.3.4 Sistemas, plataformas e estudos de caso . . . 15
2.4 Ética e desafios . . . 16

**3 Metodologia** **18**

3.1 Pesquisa documental . . . 18
3.1.1 Bases de dados . . . 19
3.1.2 Estratégia de pesquisa e *queries* . . . 19

3.1.3 Critérios de pesquisa . . . 20
3.1.4 Critérios de inclusão e exclusão . . . 20
3.1.5 Análise das fontes . . . 21
3.2 PRISMA: *Query 1* e *Query 2* . . . 21
3.3 Investigação-ação . . . 23
3.3.1 Contexto institucional: Arquivo Municipal de Viana do Castelo . . . 24
3.3.2 Caracterização da Coleção Foto Roriz . . . 24
3.3.3 Caracterização e seleção da amostra . . . 25

**4 Conceção e Desenvolvimento do Modelo de Descrição 26**
4.1 Preparação da amostra e levantamento . . . 26
4.2 Construção do vocabulário controlado . . . 29
4.3 Orientações para o guia de descrição . . . 32
4.4 Módulo de Descrição de Imagem . . . 33
4.5 Ajuste metodológico . . . 35
4.6 Teste complementar ChatGPT Vision . . . 35

**5 Resultados e Discussão 37**
5.1 Resultados da descrição humana normalizada . . . 37
5.2 Resultados do módulo de descrição de imagem . . . 39
5.3 Resultados do teste com ChatGPT Vision . . . 43
5.4 Comparação entre abordagens . . . 44
5.5 Impacto na recuperação da informação . . . 46
5.6 Discussão dos resultados . . . 47
5.7 Implicações para os arquivos . . . 48

**6 Conclusões 50**

**Referências bibliográficas 52**

**Anexos 55**
Anexo A – Lista de termos da amostra . . . 57
Anexo B – Guia para a descrição de imagens . . . 58
Anexo C – Resultados do Módulo de Descrição de Imagem . . . 65
Anexo D – Resultados do ChatGPT Vision . . . 68

# Lista de Figuras


1.1 Árvore de objetivos . . . . . 4

3.1 Diagrama PRISMA da *Query* 1 . . . . . 22
3.2 Diagrama PRISMA da *Query* 2 . . . . . 23

4.1 Grelha de análise exploratória dos retratos . . . . . 26
4.2 Grelha de análise, campos utilizados na recolha de elementos visuais . . . . . 28
4.1 Processo de evolução do termo descritivo ao termo simples . . . . . 30
4.3 Exemplo de etiquetagem em caixas delimitadoras . . . . . 31
4.4 Exemplo de etiqueta e caixa correspondente . . . . . 31
4.5 Exemplo de descrição geral onde é mencionado o atributo "claro" . . . . . 31

5.1 Resultado da etiquetagem automática (Versão A) . . . . . 40
5.2 Resultado da etiquetagem da amostra de treino (Versão B) . . . . . 40
5.3 Resultado da etiquetagem na Versão C, para a referência AFT/FR-1/04/C/23 . . . 41
5.4 Resultado da etiquetagem manual na Versão C1 . . . . . 42
5.5 Resultado de pesquisa para "mulher saia casaco" . . . . . 43
5.6 Resultado do teste com ChatGPT Vision . . . . . 44
5.7 Resultados do segundo e terceiro filtros de pesquisa . . . . . 47

B.1 Exemplo de aplicação das etiquetas . . . . . 61
B.2 Exemplo de representação de etiqueta na imagem . . . . . 61
B.3 Resultado do processo de etiquetagem na imagem . . . . . 62
B.4 Resultado da pesquisa na imagem . . . . . 62
B.5 Resultado da pesquisa visível na descrição geral . . . . . 63
B.6 Termo validado (**aceite**) . . . . . 63
B.7 Termo validado (**rejeitado**) . . . . . 63
B.8 Registo do histórico de versões do campo descrição geral . . . . . 64
C.1 Exemplo das etiquetas geradas automaticamente na Versão A . . . . . 65
C.2 Exemplo das etiquetas atribuidas pelo utilizador na Versão B . . . . . 66
C.3 Descrição de imagem no MDI, Versão C1 . . . . . 66
C.4 Etiquetas geradas pelo MDI . . . . . 66
C.5 Etiquetas criadas pelo utilizador no MDI . . . . . 67

C.6 *Print* do campo descrição geral e histórico de versões . . . . . 67
D.1 *Print* do ChatGPT Vision *input* Teste 1 . . . . . 68
D.2 *Print output* do ChatGPT Vision, Teste 1 – Fotografia 1 . . . . . 69
D.3 *Print output* do ChatGPT Vision, Teste 1 - Fotografia 2 . . . . . 70
D.4 *Print output* do ChatGPT Vision, Teste 1 - Fotografia 3 . . . . . 71
D.5 *Print input* do ChatGPT Vision, Teste 2 . . . . . 72
D.6 *Print output* do ChatGPT Vision, Teste 2 - Fotografia 1 . . . . . 73
D.7 *Print output* do ChatGPT Vision, Teste 2 - Fotografia 3 . . . . . 74
D.8 *Print output* do ChatGPT Vision, Teste 2 - Resumo . . . . . 75

# Lista de Tabelas


3.1 Perguntas de investigação e expressões de pesquisa . . . . . 20

4.1 Código de referência atribuído à amostra com base no ponto 3.1.1 da ISAD(G) . 27
4.2 Exemplo de conversão de termos compostos em termos simples . . . . . 30
4.3 Alinhamento de normalização com o *AAT Getty* . . . . . 32
4.4 Fase de teste do Módulo de Descrição de Imagem . . . . . 34

5.1 Categorias e exemplos de termos aplicados . . . . . 37
5.2 Frequência de termos na descrição humana . . . . . 38
5.3 Indicadores de esforço na descrição das imagens . . . . . 45
5.4 Comparação do número de etiquetas entre versões . . . . . 45
5.5 Demonstração de recuperação de informação . . . . . 46
5.6 Síntese de resultados . . . . . 48

A.1 Lista de termos da amostra . . . . . 57

# Abreviaturas, Siglas e Acrónimos

| | |
|---|---|
| AMVCT | Arquivo Municipal de Viana do Castelo |
| BD | Base de dados |
| CBIR | Técnica de Recuperação de Imagem Baseada em Conteúdos |
| CNN | Redes Neurais Convolucionais |
| DP | Aprendizagem profunda (*Deep Learning*) |
| FAIR | Fáceis de encontrar, Acessíveis, Interoperáveis e Reutilizáveis |
| IA | Inteligência Artificial |
| ISAD(G) | Norma Geral Internacional de Descrição Arquivística |
| MDI | Módulo de Descrição de Imagem |
| ML | Aprendizagem Automática (*Machine Learning*) |
| ODA | Orientações para a Descrição Arquivística |
| Q1 | *Query* 1 |
| Q2 | *Query* 2 |
| RI | Recuperação da Informação |
| RSL | Revisão Sistemática de Literatura |
| SI | Sistema de Informação |
| VC | Visão Computacional |

# Capítulo 1

# Introdução

## 1.1 Contextualização

O aumento da digitalização dos acervos fotográficos tem contribuído para tornar as imagens preservadas em arquivos, bibliotecas e demais instituições de memória cada vez mais acessíveis. No entanto, a recuperação dessas imagens depende de metadados capazes de representar o conteúdo e o contexto das fotografias. Num contexto desafiante, como o atual, a Inteligência Artificial (IA) surge como uma tecnologia capaz de contribuir para metadados mais ricos, conduzindo à otimização dos processos de descrição e recuperação da informação.

A presente dissertação, intitulada "A Inteligência Artificial ao serviço do Arquivo, da Identidade e da Memória", foi realizada no âmbito do mestrado em Ciência da Informação da Faculdade de Engenharia da Universidade do Porto. O estudo foi desenvolvido em ambiente institucional, mais concretamente no Município de Viana do Castelo - UO 3º Grau Arquivo e Memória, e centra-se na análise da utilização da IA na descrição de retratos de estúdio e no seu contributo para a recuperação da informação em contexto arquivístico.

O estudo foi desenvolvido com base na Coleção Foto Roriz, custodiada pelo Arquivo Municipal de Viana do Castelo, e visa analisar como a tecnologia de IA pode auxiliar na descrição de retratos de estúdio e contribuir para a produção de metadados mais ricos, consistentes e capazes de promover e potenciar novas formas de acesso e exploração da informação.

Assim, antes de apresentar o problema de investigação, é importante enquadrar a fotografia enquanto documento de arquivo e os desafios que a sua descrição impõe à recuperação da informação em ambiente digital.

A tarefa de descrever e classificar imagens de forma pormenorizada continua a ser morosa e a exigir tempo e recursos humanos quando realizada de forma manual. Por outro lado, as potencialidades da IA, em especial no que se refere ao reconhecimento de padrões visuais, à classificação automática e à extração de metadados, têm sido apontadas como soluções promissoras para enfrentar estes desafios nos arquivos e em coleções visuais. Estudos recentes destacam que os modelos baseados em *machine learning* podem automatizar elementos da recuperação de metadados e enriquecer o conteúdo descritivo, permitindo que os arquivistas se concentrem em tarefas de caráter interpretativo e analítico (Formaro e Chiquet, 2020).

Segundo Bushey (2024), a aplicação de IA, em particular a visão computacional, representa uma mudança de paradigma na gestão de arquivos fotográficos, devido ao crescimento do volume de imagens digitais e digitalizadas. Embora os métodos tradicionais de descrição continuem a ser indispensáveis, a sua aplicação exclusivamente manual pode tornar-se insuficiente, do ponto de vista operacional, quando aplicada a grandes acervos, podendo a IA contribuir para otimizar tarefas de organização, descrição, indexação e pesquisa.

De acordo com Bushey (2024), esta tecnologia é aplicada não só na extração automática de metadados, reconhecimento de objetos e atributos, reconhecimento facial, mas também na pesquisa e organização das coleções. No entanto, a autora refere que esta abordagem levanta também questões éticas muito importantes, como o viés nos dados de treino e o equilíbrio entre a automatização e a intervenção humana, que deverão ser analisadas de forma crítica ao longo de todo o projeto.

O Arquivo Municipal de Viana do Castelo (AMVCT) tem à sua guarda vários acervos fotográficos, incluindo retratos de estúdio do início do século XX, dos quais foi retirada uma amostra para esta investigação. O presente estudo vai procurar mostrar como a IA pode contribuir para melhorar os processos de descrição, classificação e recuperação da informação, respeitando as normas arquivísticas e os princípios éticos.

## 1.2 Motivação

A motivação para este projeto nasce da necessidade de tornar o AMVCT num serviço mais dinâmico, moderno, eficiente e próximo dos seus utilizadores. Os serviços de arquivo enfrentam novos desafios na gestão e divulgação do seu património cultural, neste contexto, a integração de tecnologias, como a inteligência artificial constitui uma oportunidade para modernizar as formas de trabalhar, otimizar rotinas e reforçar a valorização dos arquivos fotográficos, em particular, das coleções de retratos de estúdio.

Com esta investigação, pretende-se contribuir para a agilização de processos de descrição e indexação, através da automatização de tarefas morosas, como a classificação e a extração de metadados. Pretende-se ainda reforçar a consistência da informação associada às imagens, promovendo uma descrição mais completa, normalizada e adequada à recuperação da informação. O recurso à IA poderá tornar o trabalho técnico mais eficiente, contribuindo para benefícios com impacto direto no dia a dia do serviço. A automatização de tarefas simples libertará recursos humanos para tarefas como o atendimento, estudos e investigação.

Em simultâneo, a IA poderá ser uma mais-valia para o enriquecimento das descrições, alinhando-se com os atuais padrões de preservação digital e o cumprimento dos princípios FAIR[1] e, assim, melhorar a pesquisa interna, o acesso público à informação e a interoperabilidade entre sistemas.

Além disso, este projeto permitirá, por um lado, reforçar a missão do serviço de arquivo, com o objetivo de retirar o acervo da "caixa", atribuindo-lhe valor único, onde cada retrato seja mensageiro de identidade e memória, tornando o arquivo um serviço ainda mais aberto. Por outro lado, este projeto tem também um propósito científico: procurar compreender de que modo a IA

[1] Princípios FAIR: fáceis de encontrar, acessíveis, interoperáveis e reutilizáveis. https://allea.org/european-code-of-conduct-2017/ [consultado em 30 de novembro de 2025]

pode complementar o trabalho dos arquivistas, sem os substituir, mas contribuir para descrições mais consistentes.

## 1.3 Enquadramento do projeto

O objeto de estudo deste projeto incidirá na descrição de uma amostra de 350 retratos, a retirar da coleção Foto Roriz. Este acervo é composto por cerca de cinco mil retratos cuja origem remonta ao início do século XX, mais precisamente, 1911. A coleção tem sido reconhecida como *"um tesouro de família que passou assim a ser um tesouro da cidade"*[2], constituindo um testemunho visual de grande valor para a história social e cultural de Viana do Castelo.

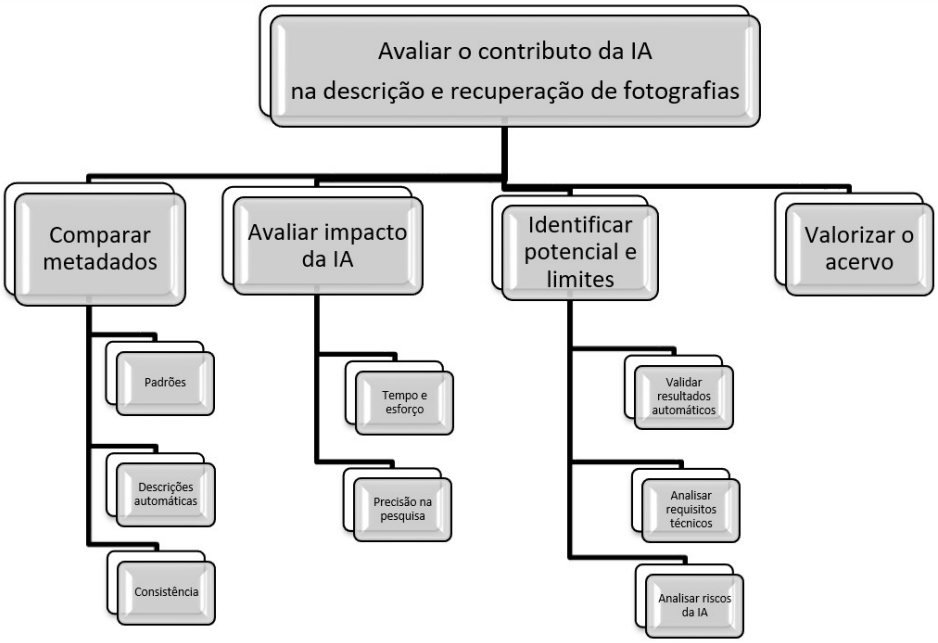
Para o estúdio Roriz, as fotografias eram mais do que simples retratos, eram gestos de dignidade e memória. Muitos dos retratados chegavam ao estúdio descalços e calçavam-se apenas para a fotografia. Era naquele espaço que as noivas eram fotografadas sozinhas, para casar sem o noivo presente, esposas e filhos com expressões sérias, destinadas a um marido ausente, e até bebés que sorriam para um pai longínquo. Cada retrato conserva o registo das ausências e dos afetos, da tradição e da *"chieira"*[3] registos do quotidiano de uma cidade feita com identidade, tradição e memória. Foi nas palavras da família Roriz, que ainda hoje tem loja aberta ao público na cidade, que nasceu a vontade de colmatar esta lacuna existente no AMVCT.

## 1.4 Objetivos

O presente projeto tem como objetivo comparar e avaliar o contributo da IA na descrição e recuperação da informação arquivística, utilizando como caso de estudo uma amostra dos retratos de estúdio que integram o acervo Foto Roriz à guarda do AMVCT. Em primeiro lugar, pretende-se comparar a eficácia de esquemas de metadados elaborados manualmente pelos técnicos do AMVCT com os esquemas de metadados gerados por meio de sistemas suportados por IA que integram o sistema de informação em uso, depois analisar o nível de consistência, profundidade e normalização das descrições. Em segundo lugar, avaliar o impacto da IA quanto à eficiência e qualidade dos processos de indexação e recuperação da informação, observando a experiência e o *feedback* dos utilizadores em diferentes contextos de pesquisa. Em paralelo, identificar as vantagens e limitações associadas à aplicação de IA nos arquivos, considerando os aspetos técnicos, éticos e epistemológicos inerentes à mediação tecnológica. Por fim, este projeto pretende valorizar o acervo fotográfico e promover o acesso à memória visual e à identidade local, por meio da difusão do património fotográfico. A Imagem 1.1 representa a árvore dos objetivos desenvolvida.

[2] Público. https://www.publico.pt/2016/01/23/local/noticia/roriz-um-estudio-onde-os-pobres-de-viana-se-calcavam-para-a-fotografia-1721154 [consultado em 30 de novembro de 2025]

[3] Termo local usado para designar vaidade, orgulho e altivez da mulher vianesa. https://www.investigo.biblioteca.uvigo.es/xmlui/handle/11093/854 [consultado em 30 de novembro de 2025]




Fonte: elaboração própria

Imagem 1.1: Árvore de objetivos

## 1.5 Problemas identificados e a resolver

Um dos principais problemas identificados prende-se com a falta de recursos humanos, pois, apesar de os técnicos possuírem formação na área de Arquivo, o número de profissionais é insuficiente para uma tarefa desta dimensão e complexidade, uma vez que o processo tradicional de descrição é demorado, depende da experiência de cada técnico, tende a introduzir subjetividade que compromete a uniformidade dos registos e dificulta a recuperação da informação.

Outro problema identificado é ao nível da descrição documental, nomeadamente ao nível do documento simples, onde se observa a repetição de títulos como "Retrato de Estúdio: casamento", "Retrato de Estúdio: família" ou "Retrato de Estúdio: casal" com referência apenas à data e ao suporte. Esta descrição dificulta a recuperação da informação mais específica, especialmente no que respeita aos aspetos identitários e culturais da tradição vianense, como o uso do ouro, a representação da mãe com os filhos ou a ausência da figura masculina, elementos que refletem contextos sociais, económicos e culturais próprios de uma época.

Além disso, embora o sistema de informação utilizado no AMVCT adote a Folha de Recolha de Dados (FRD) com base na ISAD(G)[4], os campos de descrição são de preenchimento livre, o que permite uma grande variação terminológica entre os registos. Essa flexibilidade, aliada à inexistência de um modelo normalizado de metadados adaptado ao contexto das fotografias, leva a inconsistências da informação descritiva e a problemas de interoperabilidade com outros sistemas.

Deste modo, procura-se minimizar o tempo de descrição, a subjetividade e a falta de normaliza-

[4]Norma Geral Internacional de Descrição Arquivística. https://arquivos.dglab.gov.pt/servicos/documentos-tecnicos-e-normativos/descricao/ [consultado em 30 de novembro de 2025]

ção nos processos de descrição e indexação de retratos de estúdio, com vista a uma recuperação da informação mais eficiente.

## 1.6 Estrutura da Dissertação

Além da presente introdução, esta dissertação encontra-se organizada em cinco capítulos principais. O Capítulo 2 apresenta a revisão sistemática da literatura, na qual são analisados os principais conceitos, abordagens e estudos relacionados com a temática em investigação. No Capítulo 3 é descrita a metodologia adotada, detalhando os procedimentos, métodos e técnicas utilizados para a concretização do estudo. O Capítulo 4 é dedicado à conceção e ao desenvolvimento do modelo de descrição proposto, apresentando as suas principais características e etapas de implementação. No Capítulo 5 são apresentados e discutidos os resultados obtidos, procedendo-se à sua análise crítica à luz da literatura e dos objetivos definidos. Por fim, o último capítulo reúne as conclusões do trabalho, destacando os principais contributos da investigação, as suas limitações e possíveis linhas de investigação futura.

# Capítulo 2

# Revisão Sistemática de Literatura

## 2.1 Enquadramento conceptual

Esta secção pretende facultar a base teórica necessária para compreender os pontos principais da investigação. Os referidos pontos são: o objeto de estudo (fotografia), o contexto institucional e cultural (arquivos fotográficos) e os procedimentos de organização e descrição da informação (metadados, esquemas de metadados, indexação e recuperação da informação). Esclarecer o que os conceitos querem dizer é importante para entender a pertinência da necessidade de descrever e gerir as coleções fotográficas, antes de se passar para as secções seguintes.

A abordagem metodológica utilizada no desenvolvimento desta revisão sistemática da literatura será detalhada no Capítulo 3.

### 2.1.1 Fotografia

Segundo a literatura, entende-se por fotografia uma imagem analógica, digitalizada ou nativa digital, à guarda de uma instituição, com o objetivo de ser um testemunho visual de um determinado contexto. O seu valor de "arquivo" depende de três fatores: em primeiro, da natureza e do formato, onde se inclui a fotografia tradicional, digitalizada ou digital, e tratada como objeto digital quando incorporada num sistema informático. Em segundo, a sua proveniência, pois, para que seja um documento de arquivo de valor, é necessário identificar quem o criou, o que representa, onde e quando foi produzido e com que objetivo. Por fim, em terceiro, a fotografia representa elementos do mundo real, que nos permite compreender a realidade e o contexto em que foi produzida (Bushey, 2024; Triques et al., 2024).

No entanto, o crescimento exponencial das coleções fotográficas tem dificultado uma descrição detalhada e consistente, o que leva a lacunas nos metadados, designadas por (Alkemade et al., 2023) como *Archive Gap*. Neste contexto, a fotografia deve ser entendida como um documento interpretável, pelo que é possível recorrer a ferramentas como a IA para facilitar a criação de metadados, e estes "não devem ser aceites de forma absoluta", mas sim, que possam ser validados por especialistas, de modo a criar um *ground truth* para treinar e avaliar as soluções de IA, (Triques et al., 2024; Abgaz et al., 2021).

### 2.1.2 Arquivos fotográficos

Na literatura, arquivos fotográficos são entendidos como repositórios de documentos únicos da história, funcionam como lugares de memória, onde a tradição e as evidências do passado são materializadas e armazenadas. Para compreender a natureza e a função dos arquivos, é importante ter em conta alguns elementos essenciais, como a diversidade de suportes que os constituem, podendo ser negativos de vidro, papel ou celuloide, impressões em álbuns, pastas ou livros e, mais recentemente, imagens digitais (Truyen e Fresa, 2016). Também garantir que os documentos continuem autênticos e íntegros, que as imagens possam ser recuperadas no futuro sem alterações, e que estejam protegidas de perdas ou mudanças causadas por fatores políticos, sociais ou culturais (Park, 2023).

As fotografias constituem testemunhos sociais e emocionais e são o espelho das alterações da sociedade ao longo do tempo, elas fazem com que seja possível reconstruir acontecimentos e experiências coletivas (Pierce, 2019). Por isso, conhecer a proveniência e o processo de criação das imagens é essencial, pois permite compreender a intenção do fotógrafo (Lee, 2021).

Atualmente, muitos arquivos têm vindo a afirmar-se para além de simples depósitos de imagens, assumindo-se como ecossistemas digitais inteligentes, apoiados na utilização de metadados e tecnologias digitais para melhorar a pesquisa, o acesso e a interpretação das imagens (Gowri et al., 2025). Além disso, os arquivos fotográficos ajudam a preservar a memória coletiva e histórias que poderiam ser esquecidas.

### 2.1.3 Metadados

No contexto da fotografia, os metadados não devem ser entendidos apenas como "dados sobre dados", mas como instrumentos dinâmicos de mediação, interpretação e reflexão ética, capazes de contextualizar imagens, priorizar a sua recuperação e influenciar a forma como são compreendidas pelos utilizadores, tornando-as localizáveis, acessíveis, contextualizadas e interpretáveis no âmbito de processos sociais, culturais e históricos (Pierce, 2019).

Do ponto de vista técnico e funcional, os metadados incluem informação administrativa, descritiva e estrutural, pois possibilitam interoperabilidade entre sistemas e repositórios digitais, como é o caso da Europeana[1] e da DPLA (Digital Public Library of America[2]). Para isso, utilizam esquemas e normas como o EDM[3] (Europeana Data Model) ou o MAP[4] (Metadata Application Profile), padrões que permitem que diferentes repositórios comuniquem entre si (Truyen e Fresa, 2016). Enquanto pontos de acesso, os metadados permitem pesquisar e recuperar coleções em larga escala por título, data, assunto ou outros descritores. Além de cumprirem funções técnicas, os metadados também permitem preservar a memória e a identidade, na medida em que transformam registos estáticos em histórias digitais que reforçam a identidade da comunidade ao mesmo tempo que permitem a conexão entre objetos de diferentes repositórios (Davidson e Reid, 2022). Ao

[1] Europeana: https://www.europeana.eu/pt
[2] DPLA: https://dp.la/
[3] EDM: https://pro.europeana.eu/page/edm-documentation
[4] MAP: https://pro.dp.la/hubs/metadata-application-profile

utilizar tesauros ligados, como o AAT[5] (*Art and Architecture Thesaurus*) ou o SKOS[6] (*Simple Knowledge Organization System*), permite que os metadados representem aspetos mais específicos das imagens, como estilos, técnicas ou contextos de utilização.

As decisões humanas e os algoritmos influenciam os metadados, que podem revelar ou ocultar informação e refletir escolhas culturais, históricas e institucionais do arquivo (Lee, 2021). Do ponto de vista ético, os metadados são um espaço de responsabilidade que garante a dignidade das pessoas representadas, mas também ajudam a corrigir desequilíbrios históricos de poder, o que é especialmente importante em arquivos sensíveis ou ligados a comunidades marginalizadas Hrechaniuk e Sparrman, 2024). Além da função técnica, os metadados exercem uma função dinâmica, ao registar decisões realizadas durante a digitalização e classificação das imagens. Conhecido como paradados, este processo incorpora contribuições do público e conteúdos produzidos pelos utilizadores, o que melhora a consulta e a contextualização dos arquivos (Late et al., 2024).

### 2.1.4 Esquemas de metadados

Um esquema de metadados é entendido como uma estrutura lógica e formal criada para organizar, gerir e descrever informação sobre objetos digitais ou físicos (Lo Turco et al., 2019). Nos sistemas de bases de dados relacionais, corresponde ao modelo de organização da informação, que define a relação entre os objetos e os atributos que os descrevem. As funções essenciais para esquemas de metadados tendem a garantir a normalização e a interoperabilidade, permitindo que diferentes sistemas e instituições consigam comunicar entre si.

Plataformas como o DPLA, *Metadata Application Profile* (MAP) e o *Europeana Data Model* (EDM) funcionam como um padrão, que reúne dados de milhares de instituições distintas (Truyen e Fresa, 2016; Bachi et al., 2018). Um esquema bem estruturado de dados organiza os metadados em diferentes categorias: descritivos, que registam o conteúdo intelectual e visual do objeto; administrativos, que documentam a gestão e a conservação; e estruturais, que descrevem as relações entre as partes de um objeto. Atualmente, já existem esquemas de metadados padronizados como o *Dublin Core*[7], que fornece interoperabilidade para metadados descritivos e técnicos, o *Cataloging Cultural Objects*[8] (CCO), que abrange metadados descritivos e administrativos e serve de modelo para arquivos de museus e igrejas, os esquemas EXIF (*Exchangeable Image File Format*) e IPTC (*International Press Telecommunications Council*), que são específicos para fotografia e registam, respetivamente, definições técnicas da câmara e gestão de ativos de multimédia e o EAD (*Encoded Archival Description*), um padrão de descrição para arquivos.

A definição de padrões é fundamental para a gestão e computação dos dados, mas os autores alertam que esquemas pouco flexíveis podem não capturar toda a complexidade ou dimensão simbólica de certos objetos históricos. Um esquema de metadados atua como estrutura central de uma coleção, ele transforma arquivos passivos em ecossistemas digitais ativos, inteligentes e pesquisáveis, ao fornecer os pontos de acesso necessários para a recuperação, interpretação e contextualização da informação. Desta forma, o património cultural pode ser estudado, reutilizado

[5] AAT: https://www.getty.edu/research/tools/vocabularies/aat/
[6] SKOS: https://www.w3.org/2004/02/skos/
[7] https://www.dublincore.org/
[8] https://www.vraweb.org/cco

e partilhado de forma consistente e confiável (Lo Turco et al., 2019; Truyen e Fresa, 2016).

### 2.1.5 Indexação

A indexação pode ser entendida como o processo de atribuir palavras-chave, etiquetas ou metadados descritivos a recursos, para facilitar a sua organização, pesquisa e recuperação num sistema de informação. A indexação é a tradução do conteúdo visual em descrições acessíveis a todos, assim, é possível procurar uma fotografia da mesma forma que procura um documento de texto. Trata-se de um processo que organiza o conhecimento, cria padrões do conteúdo e do contexto das imagens descritas. Assim, o sistema passa a entender o que a imagem mostra e de que forma (Adnan et al., 2022; Triques et al., 2024).

Nos serviços de arquivo, a indexação tem um papel fundamental no preenchimento de lacunas de informação conhecidas por *Archive Gap*. O termo é utilizado quando a descrição do conteúdo visual e da origem das imagens é pouco detalhada (Alkemade et al., 2023; Bushey, 2024).

Devido ao crescimento rápido das coleções digitais, a indexação realizada de forma exclusivamente manual pode enfrentar desafios de escala. Considerando esta realidade, o recorrer a tecnologias de *Deep Learning*, pode apoiar na automatização de extração de características visuais e classificação das fotografias (Bushey, 2024; Adnan el al., 2022; Eiler et al., 2018). Deste modo, procura-se garantir maior consistência e interoperabilidade no processo de descrição recorrendo a vocabulários controlados (*Linked Open Data - LOD*) e ontologias garantindo uma base semântica, como referem (Alkemade et al., 2023).

Indexar significa transformar informação visual e contextual das imagens em metadados estruturados com o objetivo de facilitar o acesso, a pesquisa e a investigação ao património cultural (Abgaz et al., 2021).

### 2.1.6 Recuperação da informação

Pode-se entender recuperação da informação (RI) como o processo de localizar e aceder a conteúdos de interesse para o investigador, num universo muitas vezes "incalculável", tão grande que chega a ser impossível de gerir manualmente. O processo está dependente da configuração do sistema e dos objetos de informação, ambos influenciados pelo modo como o investigador pensa e procura a informação.

No que toca ao contexto das imagens, os métodos de acesso podem basear-se na técnica de Recuperação de Imagem Baseada em Conteúdo (RIBC), que utiliza características de baixo nível, como, por exemplo, a cor, a forma e a textura. Este método recupera imagens a partir de atributos visuais, mas, em alternativa, também podem recorrer a *tags* ou palavras-chave que traduzem o conteúdo visual em descrições semânticas, permitindo a recuperação da imagem como se fosse texto (Adnan et al., 2022; Triques et al., 2024).

Por sua vez, o paradigma da IA introduziu a visão computacional, que permitiu que as máquinas fossem capazes de interpretar dados parecidos com a visão humana, permitindo aos investigadores realizarem pesquisas diretamente a partir das imagens, nas quais o conteúdo visual é categorizado e consultado de forma automática, devolvendo ao utilizador resultados rápidos, relevantes e personalizados da informação (Eiler et al., 2018).

No sentido de melhorar a recuperação da informação, a literatura recomenda a utilização de polirrepresentações e contexto, que combinam diferentes métodos de descrição e representação que permitem que as imagens sejam interpretadas em diferentes níveis de interação. Este processo permite enriquecer semanticamente as imagens com metadados completos sobre quem as fez, quando, onde e porquê, transformando-as em recursos científicos interpretáveis, para que o investigador possa analisar, contextualizar e utilizar as imagens nos seus estudos (Triques et al., 2024; Abgaz et al., 2021).

Ainda assim, a RI também enfrenta desafios, devido à falta de metadados detalhados em arquivos fotográficos que geram lacunas de informação. Para colmatar estas lacunas, pode justificar-se recorrer a modelos de *Deep Learning* para anotação automática, capazes de preencher as lacunas deixadas pela descrição manual e facilitar a organização e recuperação de grandes coleções, que seriam impossíveis de processar manualmente (Adnan et al., 2022).

## 2.2 Metadados em fotografias

A presente subsecção foca-se na aplicação de metadados em fotografias no contexto do património cultural, considerando o seu papel na organização, descrição, recuperação e reutilização de imagens em ambientes digitais. A literatura aponta que os metadados são fundamentais para facilitar o acesso e a pesquisa de coleções fotográficas, além de assegurarem a preservação do contexto e da proveniência dos documentos visuais (Truyen e Fresa, 2016; Late, Ruotsalainen e Kumpulainen, 2024).

A subsecção é organizada em quatro eixos principais. Numa primeira fase, analisa-se o papel dos metadados em fotografias de arquivo, com destaque para o seu contributo na mediação entre as imagens, os sistemas de informação e os diferentes perfis de utilizadores, bem como para a preservação do contexto e da proveniência dos documentos fotográficos (ierce, 2019; Davidson e Reid, 2022P). Na segunda fase, aborda-se o papel dos metadados descritivos, enquanto instrumentos centrais para a representação do conteúdo visual e contextual das fotografias, fundamental para a indexação e a recuperação da informação em arquivos e repositórios digitais (Park, 2023). Na terceira fase, procede-se à apresentação dos principais esquemas e normas de metadados aplicados às fotografias, com particular atenção às iniciativas de normalização e interoperabilidade no domínio do património cultural digital, incluindo projetos colaborativos de âmbito europeu (Bachi et al., 2018; Truyen e Fresa, 2016). Por fim, discutem-se as limitações da descrição arquivística tradicional, nomeadamente no que respeita à subjetividade interpretativa, à inconsistência descritiva e à dificuldade de escalar a descrição manual face ao crescimento das coleções fotográficas digitais (Pierce, 2019; Alkemade et al., 2023).

### 2.2.1 Funções dos metadados em fotografias

Os metadados desempenham um papel central na gestão, preservação, acesso e reutilização das fotografias de arquivo, em especial no contexto do património cultural digital. De acordo com a literatura, as imagens fotográficas, por si só, não são suficientemente acessíveis nem recuperáveis sem uma camada descritiva que lhes atribua contexto, significado e relações (Bachi et al., 2018;

Sarkar e Biswas, 2020).

No âmbito de grandes agregadores e infraestruturas digitais, como a Europeana ou a *Digital Public Library of America*, os metadados permitem não só a pesquisa e a interoperabilidade entre coleções heterogéneas, mas também a sua integração em ecossistemas mais amplos de dados culturais (De Jager-Loftus, 2024; Wolff et al., 2024). Esta função é particularmente relevante em arquivos fotográficos históricos, onde os objetos visuais são frequentemente reutilizados para fins científicos, educativos e criativos.

Também Late et al, (2024) mostram que a eficácia da recuperação de imagens em arquivos fotográficos está diretamente relacionada com a qualidade, consistência e granularidade dos metadados disponíveis, influenciando as estratégias de pesquisa adotadas pelos utilizadores e os obstáculos enfrentados durante a investigação histórica.

### 2.2.2 Metadados descritivos

Os metadados descritivos são a principal camada que permite que a fotografia de arquivo seja entendida pelo utilizador. Têm um papel central na identificação, contextualização e interpretação das imagens em sistemas de informação patrimonial e possuem informação importante como: o título, autor, data, local, assunto, tipo de fotografia e contexto de produção, o que permite não só recuperar a imagem, mas também entender o que ela representa num determinado momento histórico, social e cultural (Lo Turco et al., 2019; Pierce, 2019). No domínio da arquivística, a descrição descritiva tem uma função que interpreta, na medida em que traduz visualmente conteúdos complexos para estruturas de texto normalizadas.

A literatura refere que os metadados descritivos das fotografias de arquivo têm muitas vezes informação incompleta, imprecisa e baseada em acontecimentos passados. Muitas vezes, isto acontece porque os registos originais não existem ou estão fragmentados, devido à informação se ter perdido com o tempo, ou porque é preciso descrever as imagens depois de as criar. Pierce, (2019) mostra como estas condições influenciam diretamente a forma como as fotografias são digitalizadas, ele mostra que grupos sociais que foram sempre deixados de lado, como as mulheres, continuam a ser representados de forma a não darem visibilidade ou a serem descritos de forma geral, o que faz com que as fotografias não sejam facilmente encontradas, que não se saiba bem o que elas mostram e que não se consiga entender as coleções fotográficas.

Perante estas limitações, os autores defendem a adoção de abordagens participativas como estratégia para enriquecer e complementar os metadados descritivos. Hansson e Dahlgren, (2022), no contexto do *crowdsourcing*[9] em arquivos municipais, mostram que a participação do público pode contribuir com conhecimentos de contexto, memórias locais e interpretações alternativas que dificilmente seriam captadas por descrições institucionais tradicionais. De forma semelhante, (Davidson e Reid, 2022) sublinham que processos colaborativos de descrição permitem ampliar o significado das imagens, reforçando a ligação entre os arquivos e as comunidades a que os documentos dizem respeito.

[9]Modelo de produção e de estruturação de processos que recorre à participação de uma comunidade.

### 2.2.3 Esquemas e normas aplicadas às fotografias

A aplicação de esquemas e normas de metadados é muito importante na descrição de fotografias de arquivo, pois garante a coerência semântica, interoperabilidade entre sistemas e sustentabilidade a longo prazo da informação. As coleções fotográficas são muito diferentes em termos de material, tempo e instituições, é preciso utilizar modelos normalizados para evitar descrições fragmentadas e difíceis de reutilizar, como referem (Lo Turco et al., 2019).

Modelos de dados como o *Dublin Core*, o CIDOC CRM[10] e outros perfis específicos para o património cultural, têm sido discutidos como formas de organizar a informação descritiva, contextual e relacional associada às fotografias. Lo Turco et al.,(2019) salientam que estes esquemas permitem mostrar não só características básicas da imagem, mas também relações complexas entre objetos, pessoas, eventos e o contexto em que foram produzidos, relações que são fundamentais para a interpretação arquivística e museológica.

Iniciativas colaborativas como o PHOTOCONSORTIUM[11] mostram a importância de alinhar práticas descritivas entre instituições que têm património fotográfico, enquanto Truyen e Fresa, 2016; Bachi et al. 2018) mostraram que juntar e partilhar esquemas de metadados permite juntar e partilhar fotografias históricas em toda a Europa e, desta forma, torna mais fácil o acesso a coleções de fotografias que estão dispersas e reforça a visibilidade do património fotográfico. A experiência destes projetos mostra que os esquemas de metadados não são só instrumentos técnicos de descrição, mas também ajudam na cooperação entre instituições, na negociação semântica e na criação de infraestruturas comuns de conhecimento.

A discussão sobre normas de metadados tem sido ampliada para incluir preocupações relacionadas com a qualidade, transparência e ética dos dados, sobretudo em contextos de reutilização computacional. Alkemade et al.,(2023) sugerem que as *datasheets* para conjuntos de dados digitais de património cultural sejam utilizadas como um documento adicional aos esquemas tradicionais de metadados. Estas fichas técnicas explicam como os conjuntos de dados fotográficos foram criados, selecionados, digitalizados e organizados, bem como as suas lacunas, limitações e possíveis enviesamentos.

### 2.2.4 Limitações da descrição tradicional

Apesar de estarem centrados nos arquivos e repositórios, a descrição tradicional baseada apenas em metadados manuais tem limitações que afetam a sua eficácia e compreensão. A literatura mostra que há vários problemas: as descrições não são muito grandes, especialmente quando se trata de coleções muito grandes, as pessoas que fazem a descrição podem ter opiniões diferentes e isso pode fazer com que as interpretações também sejam diferentes, as palavras podem ser diferentes dependendo da pessoa, da instituição ou do país, e a qualidade da descrição depende muito do conhecimento e da experiência da pessoa que faz a descrição (Sarkar e Biswas, (2020); Park, 2023).

Além das questões técnicas e metodológicas, os investigadores Odumosu (2020) e Hrechaniuk e Sparrman (2024) destacam a dimensão ética da descrição arquivística. Descrições que não têm a ver com a realidade ou que não são sensíveis, principalmente em arquivos coloniais ou que têm a ver

[10] CIDOC-CRM: https://cidoc-crm.org/
[11] PHOTOCONSORTIUM: https://www.photoconsortium.net/

com pessoas que foram marginalizadas, podem fazer com que se fale menos de um assunto, reforçar estereótipos e fortalecer relações de poder implícitas. A produção de metadados não é apenas uma questão de precisão factual, mas também de responsabilidade ética. Por isso, é necessária uma reflexão crítica sobre o contexto histórico, social e cultural das fontes.

Lee (2021) complementa esta análise ao propor o conceito de "arqueologia de dados", o conceito destaca que, em muitos casos, conjuntos fotográficos muito grandes escondem camadas complexas de "mediação composta". O autor diz que as decisões mais importantes no processo de digitalização, como o que fotografar, como construir os dados de treino para algoritmos e como adotar termos classificatórios, não são visíveis nos metadados finais. Esta falta de clareza mostra que os documentos tradicionais mostram apenas uma parte da história e do que as coleções pretendiam dizer. Existe o risco de que os algoritmos sejam tendenciosos, o que pode fazer com que as narrativas históricas sejam ignoradas ou apagadas, podendo reescrever o arquivo de forma subtil, reforçando as desigualdades ou os silêncios históricos que já existiam.

## 2.3 Inteligência Artificial: automatização de metadados

A possibilidade de usar IA para automatizar a extração de metadados em arquivos fotográficos é um tema muito discutido na literatura. Isso acontece porque a descrição tradicional de arquivos tem limitações. O principal problema é o número de fotografias, tanto digitais quanto digitalizadas, que está a aumentar rapidamente. Isso torna difícil fazer uma descrição manual de cada fotografia e compromete a escalabilidade, a consistência e a precisão dos metadados (Bushey, 2024; Triques et al., 2024)).

A IA não substitui o trabalho dos arquivistas, mas é uma ferramenta que ajuda a automatizar tarefas repetitivas. Ela identifica padrões visuais e gera metadados que podem ser validados e melhorados por humanos (Bushey, 2024).

### 2.3.1 Inteligência Artificial nos Arquivos

A literatura define que a IA é a capacidade que os computadores têm de executar tarefas que, por norma, exigem inteligência humana, como, por exemplo, o reconhecimento visual, a classificação e a inferência semântica. Que, no caso dos arquivos fotográficos, se consegue, sobretudo, com a aprendizagem automática (*machine learning*) e, em particular, com a aprendizagem profunda (*deep learning*), que permite treinar modelos complexos a partir de grandes volumes de dados visuais (Câmara et al., 2023).

A visão computacional é uma parte importante deste conceito, pois é a parte da IA que interpreta automaticamente o que vemos nas imagens. Segundo Eiler et al. (2018), técnicas como as Redes Neurais Convolucionais (CNNs) permitem extrair características visuais hierárquicas diretamente dos pixels, o que significa que não é necessário definir manualmente descritores, e é por meio desta forma que é aumentada a capacidade dos modelos de se adaptarem à diversidade das fotografias que existem nos arquivos históricos.

Em relação aos arquivos, a maioria dos autores está de acordo quando referem que a IA deve ser utilizada para analisar e descrever imagens, mas isto só deve ser feito se cumprir com princípios da

arquivística e sem que a tecnologia seja o mais importante. Bushey (2024) diz que, se não tivermos cuidado, podemos usar as ferramentas de IA para analisar fotografias de uma forma errada, pois os algoritmos só mostram alguns padrões visuais e esquecem o que aconteceu antes e depois das fotografias serem tiradas. Para o autor, os arquivistas devem continuar a questionar sobre o que está representado nas imagens, por que razão, quem as produziu e em que circunstâncias, pois, se não o fizerem, a automatização pode reforçar estereótipos históricos ou narrativas dominantes, incluindo perspetivas parciais ou limitativas.

Neste caso, a regra da proveniência é a mais importante. Para Bushey (2024), a descrição arquivística serve para mostrar de onde vêm os registos visuais, a sua origem institucional, social e funcional, assim a informação não pode ser apenas inferida por meio de uma análise visual automática. Também Abgaz et al. (2021) reforçam esta ideia ao dizerem que muitas imagens digitais de património não têm informação importante sobre o que significam culturalmente, simbolicamente e historicamente. Para os autores, a IA pode enriquecer a descrição, mas só se estiver integrada em modelos que preservem e expliquem as relações de origem e de contexto.

Também a questão da autenticidade e da interpretação é destacada por Abgaz et al. (2021), que dizem que o objetivo da aplicação da IA nos arquivos não deve ser uma simples rotulagem automática, mas sim a transformação das imagens em recursos científicos interpretáveis, capazes de apoiar a investigação histórica e cultural. O que significa que é preciso entender a informação social, económica e política, sem fazer interpretações erradas ou simplistas, a literatura refere que a IA deve ser utilizada para apoiar a mediação arquivística, mas que precisa de confirmação humana para garantir que os metadados gerados respeitem o significado e o valor das imagens.

### 2.3.2 Classificação de imagem e automatização da extração de metadados

As redes neurais convolucionais (CNNs) são utilizadas para automatizar a extração de metadados, que são muito eficazes na classificação de imagens. Eiler et al. (2018) mostraram que as arquiteturas *Inception v3* e *v4* podem classificar fotografias em categorias importantes para arquivos, como retratos, paisagens, arquitetura e cenários urbanos. Isso ajuda a criar metadados automaticamente sobre o tipo de fotografia. A "aprendizagem por transferência"é uma estratégia importante para lidar com a falta de variedade nos dados históricos. Ao utilizar modelos pré-treinados em grandes bases de dados atuais, é possível ter um bom desempenho, mesmo com conjuntos de treino pequenos dedicados ao património (Eiler et al., 2018; Wu, 2024).

Outros métodos incluem o *ResNet50*[12] de Abgaz et al. (2021), o *Xception*[13] e o *EfficientNet*[14] para detectar características humanas em fotografias antigas.

Ortiz Pablo et al. (2023) mostram que esses modelos só funcionam bem se a qualidade da imagem for boa, se as técnicas de digitalização forem adequadas e se houver um desfasamento temporal entre os dados de treino e os de aplicação. Esses requisitos são muito importantes em arquivos fotográficos, pois para cada situação, surgem soluções personalizadas. Sasithradevi et al. (2024) apresentam a *MonuNet*, uma arquitetura profunda que utiliza *Spatial Attention e Channel Attention* para classificar monumentos históricos. O estudo revela o potencial da IA para extrair

[12] Rede neural profunda.
[13] Arquitetura de rede neural.
[14] Família de redes neurais.

metadados iconográficos e temáticos em coleções de património arquitetónico.

### 2.3.3 Integração de IA com esquemas de metadados

Além da classificação automática, a literatura também destaca a importância de juntar os resultados da IA a esquemas de metadados e modelos semânticos já estabelecidos. Foi em 2021 que Abgaz et al. (2021) sugeriram uma forma de fazer várias coisas em simultâneo, uma forma que utiliza Visão Computacional e tecnologias de Web Semântica, que permitem associar anotações automáticas a ontologias e modelos de *Linked Open Data*.

Esta integração liga os metadados criados por IA a vocabulários muito utilizados no património cultural, como o *Getty Art e Architecture Thesaurus* e o *Iconclass*[15]. Assim, é possível ter interoperabilidade semântica e reutilizar a informação (Abgaz et al., 2021; Triques et al., 2024).

Alkemade et al. (2023) reforçam esta ideia ao dizerem que os sistemas baseados em IA no património cultural funcionam bem quando os algoritmos são de boa qualidade e quando estão ligados a modelos conceptuais, ontologias e infraestruturas de dados que já existem. Os autores defendem que a integração semântica é essencial para garantir que os metadados automáticos são transparentes, contextualizados e interpretáveis, bem como para apoiar a sua validação e utilização por profissionais da informação.

O conceito de "polirrepresentação"apresentado por Triques et al. (2024) no âmbito do projeto piloto de classificação de imagens da Europeana defende que os metadados automáticos devem coexistir com a descrição profissional e com a contribuição de várias pessoas, formando camadas complementares de representação. Também Bushey (2024) concorda com esta ideia e diz que os metadados criados pela IA são úteis, mas não substituem a descrição humana, principalmente quando se trata de cada item.

### 2.3.4 Sistemas, plataformas e estudos de caso

Para automatizar os metadados em arquivos fotográficos, é necessário ter a infraestrutura técnica adequada. Wu (2024) mostra que a computação em nuvem é muito importante para processar imagens de alta resolução e grandes quantidades de dados, pois permite analisar automaticamente retratos históricos e encontrar padrões visuais complexos.

Os resultados dos estudos de caso mostram que cada vez mais se juntam sistemas de IA a plataformas de gestão de património e combinam-se processamento distribuído, modelos de *deep learning* e validação humana.

Câmara et al. (2023) apresentam uma visão geral da evolução destas tecnologias, com destaque para a mudança gradual dos métodos tradicionais de extração manual de metadados para abordagens baseadas em aprendizagem profunda (*deep learning*). Estas últimas são mais resistentes, podem ser escaláveis e adaptadas a diferentes contextos patrimoniais. Segundo os autores, esta mudança tecnológica permite responder melhor a objetivos complexos, como, por exemplo, a possível transformação dos arquivos visuais estáticos em recursos científicos dinâmicos e, assim, os recursos podem ser analisados, interligados semanticamente e reutilizados em vários contextos de investigação.

[15] https://iconclass.org/

## 2.4 Ética e desafios

Recorrer à IA para a extração automática de metadados e para a classificação de fotografias de arquivo, embora traga benefícios ou vantagens, também pode levantar problemas importantes do ponto de vista ético, legal e técnico. Recentemente, no artigo de Oru (2025), a autora identificou um conflito entre dois objetivos: o de tornar as coleções das instituições de património cultural acessíveis ao público, como forma de promover o conhecimento, a preservação e o acesso público para fins culturais e o de utilizar as coleções para treino de algoritmos de IA, sem controlo, sem contexto e por vezes sem consentimento para utilizar conteúdos para treino.

É neste contexto que as coleções, quando se encontram em domínio público e livres de restrições de direitos de autor, se tornam alvos privilegiados para os profissionais da IA, uma vez que, de acordo com as políticas de digitalização e acessibilidade *online* do património cultural da União Europeia (EU), podem ser utilizadas livremente. Oruc (2025) alerta que nem todo património pode ser tratado da mesma forma, especialmente quando envolve grupos vulneráveis.

Materiais sensíveis que pertencem a comunidades específicas, nomeadamente povos indígenas, podem incluir conteúdos secretos ou culturalmente reservados e quando utilizados sem autorização ou consulta prévia, podem resultar no desrespeito pelos seus direitos culturais e no reforço de desigualdades do passado. A autora destaca que certos tipos de património exigem um cuidado especial, por exemplo, materiais de comunidades específicas, incluindo objetos sagrados ou secretos de povos indígenas, pois podem ver os seus direitos comprometidos se utilizados sem consulta prévia, perpetuando desequilíbrios históricos.

Se, por um lado, a autora apresenta esta ideia, por lado, relembra que o património dissonante, relacionado com eventos negativos como colonialismo, escravatura ou conflitos armados, pode gerar resultados tendenciosos se treinado sem contexto adequado, influenciando a perceção do público sobre história e política. Também, o património dissonante, quando utilizado sem contexto, pode gerar interpretações enviesadas e influenciar negativamente a forma como o público entende a história. As limitações técnicas de IA são outro fator identificado pela autora. Os algoritmos podem ter dificuldade em interpretar nuances históricas e culturais complexas e, considerando que a aplicação da técnica *machine unlearnig*, técnica que permite eliminar dados "problemáticos" de um modelo de treino, é difícil, torna-se fundamental uma seleção cautelosa dos dados de treino de modo a precaver resultados enviesados.

Apesar de já existirem leis e diretivas da EU (Lei da UE sobre IA[16] e a Diretiva de Direitos de Autor no Mercado Único Digital[17]) que obrigam a que os dados utilizados para treinar IA sejam transparentes e que os profissionais da IA devem indicar de onde vêm os dados e como os utilizaram. No entanto, parece que as orientações não são esclarecedoras quanto ao "apagar" ou retirar os objetos correspondentes a património cultural sensível que foi indevidamente absorvido pelo modelo, identificado por Oruc (2025) como um *gap* na legislação. No artigo, recomenda-se uma abordagem colaborativa entre arquivos, as comunidades detentoras dos ativos e os profissionais de IA, com o objetivo de reduzir os riscos associados. Sugere-se ainda a utilização de *tags* que

[16] https://www.europarl.europa.eu/topics/pt/article/20230601STO93804/lei-da-ue-sobre-ia-primeira-regulamentacao-de-inteligencia-artificial

[17] https://eur-lex.europa.eu/PT/legal-content/summary/copyright-and-related-rights-in-the-digital-single-market.html

identifiquem a proveniência e restrições dos dados, designadas por *TK Labels*, de forma a garantir uma utilização ética do património cultural.

Além disso, propõe-se a definição de diferentes níveis de acesso, de modo a criar uma hierarquia de permissões que proteja os dados mais sensíveis e, assim, garantir que nem todos os utilizadores possam aceder a todos os conteúdos.

Apesar da importância da digitalização e da automatização para preservar e tornar o património acessível, Oruc (2025) defende que é necessário equilibrar a eficiência tecnológica com considerações éticas e históricas, sem priorizar apenas a tecnologia. É fundamental manter o sentido original dos recursos digitais e assegurar que os investigadores consigam compreender o seu contexto.

# Capítulo 3

# Metodologia

O presente capítulo descreve e fundamenta as opções metodológicas adotadas no desenvolvimento da dissertação. A metodologia assume um papel central na investigação, na medida em que define os procedimentos utilizados para a construção do enquadramento teórico e para a análise empírica realizada, garantindo rigor científico, coerência metodológica e transparência no processo investigativo.

Tendo em conta os objetivos do estudo e a sua natureza teórico-prática, a metodologia organizou-se em duas vertentes complementares. A primeira corresponde à pesquisa documental, que sustentou a elaboração da revisão de literatura. Nesta fase, descrevem-se os métodos de pesquisa adotados, as bases de dados científicas consultadas, as estratégias de pesquisa utilizadas, bem como os critérios de inclusão, exclusão e análise das fontes, de modo a assegurar a relevância e a qualidade da informação selecionada.

A segunda vertente metodológica refere-se ao desenvolvimento da investigação em contexto institucional, concretamente por meio de uma investigação-ação a realizar no Arquivo do Município de Viana do Castelo. Neste subcapítulo são apresentados os métodos e técnicas de recolha e tratamento da informação aplicados em ambiente arquivístico, os procedimentos de análise adotados e as especificidades do contexto institucional que enquadram o desenvolvimento do trabalho.

Esta estrutura metodológica permite distinguir claramente a componente de fundamentação teórica da componente empírica da investigação, evidenciando a sua articulação no âmbito da Ciência da Informação e a sua contribuição para a consecução dos objetivos da dissertação.

## 3.1 Pesquisa documental

Nesta primeira vertente, é descrita a metodologia de investigação documental adotada, que inclui as bases de dados consultadas, as estratégias de pesquisa utilizadas e os critérios de seleção e análise das fontes de informação. O objetivo foi obter informação relevante e de qualidade, que permitisse realizar uma revisão da literatura sobre metadados em fotografias de arquivo e sobre a aplicação de inteligência artificial na extração de metadados e na classificação de fotografias de arquivo.

### 3.1.1 Bases de dados

A presente pesquisa documental foi realizada em três bases de dados: a *Scopus*[1], a *Web of Science*[2] e o Catálogo da Universidade do Porto (UP). As duas primeiras garantem acesso a publicações científicas de elevada qualidade e permitem aplicar diferentes filtros para uma revisão sistemática e excluir resultados irrelevantes. No que toca ao Catálogo da UP, foram consultados apenas recursos online, desde artigos até livros, e teses e dissertações, para garantir o acesso a conteúdos digitais credíveis e alinhados com a investigação.

Em todas as bases de dados, a pesquisa foi realizada com recurso aos critérios de pesquisa "Título do artigo", "Resumo" e "Palavras-chave", com o objetivo de recuperar fontes de informação relacionadas com o tema da investigação.

### 3.1.2 Estratégia de pesquisa e *queries*

No contexto da estratégia de pesquisa, foi realizado um levantamento prévio de todas as palavras-chave relevantes para o tema da investigação.

Desta análise obtiveram-se os seguintes termos: *Archives, Artificial intelligence, Automatic metadata, Cultural Heritage, Descriptive metadata, IA, Image classification, Indexing, Machine Learning, Metadata, Photographic Archives, Photographic metadata, Photographs, Photography, Studio portraits*.

A partir do conjunto de palavras-chave, procurou-se, numa fase inicial, construir uma única expressão de pesquisa que fosse capaz de incluir todos os termos relevantes. Contudo, após uma primeira análise, concluiu-se que seria mais eficaz desenvolver duas expressões de pesquisa distintas, orientadas para temas específicos: uma dedicada a metadados e outra mais centrada na inteligência artificial.

Esta decisão teve por base o facto de os estudos sobre metadados apresentarem um desenvolvimento histórico mais prolongado do que os estudos sobre inteligência artificial aplicada a contextos arquivísticos, pelo que a utilização de uma única expressão de pesquisa poderia revelar-se limitadora. Optou-se, assim, pela definição de duas expressões de pesquisa autónomas.

Para construir as expressões de pesquisa, recorreu-se aos operadores booleanos *"AND"* e *"OR"*, de modo a combinar sinónimos, siglas e conceitos relacionados, e assim recuperar o maior número possível de fontes de informação relevantes.

A seguir, na Tabela 3.1, são apresentadas as expressões de pesquisa construídas para responder às perguntas de investigação.

[1] https://www.scopus.com/
[2] https://www.webofscience.com/

Tabela 3.1: Perguntas de investigação e expressões de pesquisa

| **Pergunta de investigação** | **Expressão de pesquisa** |
|---|---|
| Q1: Como se aplicam metadados em fotografias? | *(“METADATA” OR “DESCRIPTIVE METADATA” OR “INDEXING” OR “PHOTOGRAPHIC METADATA”) AND (“PHOTOGRAPHS” OR “PHOTOGRAPHY” OR “STUDIO PORTRAITS” OR “PHOTOGRAPHIC ARCHIVES”) AND “CULTURAL HERITAGE”)* |
| Q2: Como pode a IA automatizar a extração de metadados em arquivos fotográficos? | *(“ARTIFICIAL INTELLIGENCE” OR “IA” OR “MACHINE LEARNING”) AND (“AUTOMATIC METADATA” OR “IMAGE CLASSIFICATION”) AND (“PHOTOGRAPHS” OR “PHOTOGRAPHIC ARCHIVES”) AND (“CULTURAL HERITAGE” OR “ARCHIVES”)* |

### 3.1.3 Critérios de pesquisa

Foram definidos critérios de pesquisa para todas as bases de dados (BD) analisadas, para garantir consistência, rigor e precisão na recuperação da informação. A pesquisa incidiu sobre os campos do título, resumo e palavras-chave, porque se considera que estes elementos resumem de forma mais direta o conteúdo, o tema e os objetivos dos estudos, e permitem uma seleção inicial mais eficaz e alinhada com as perguntas de investigação.

Em relação ao período temporal, foram definidos intervalos diferentes para cada pergunta de investigação. Para a questão de investigação Q1, que se refere aos metadados em fotografias de arquivo, considerou-se um período entre 2015 e 2025. Já para a questão de investigação Q2, que se refere à aplicação da inteligência artificial para extrair automaticamente metadados, o intervalo inicial foi de 2020 a 2025, por ser uma área de investigação mais recente e em desenvolvimento, no entanto, foi necessário alargar o período temporal para 2018 a 2025 de modo a aumentar o número de resultados.

Em relação aos estudos, foram incluídos apenas os que estavam no idioma inglês, esta opção deve-se ao inglês ser a principal língua de comunicação na ciência e ao acesso diversificado a fontes académicas.

Por último, foi aplicado o critério Acesso Aberto, para garantir que os documentos analisados estavam totalmente disponíveis para consulta e verificação.

### 3.1.4 Critérios de inclusão e exclusão

A revisão de literatura seguiu critérios específicos para garantir que apenas os estudos diretamente relevantes ao tema fossem considerados.

Assim, foram incluídos estudos sobre como utilizar metadados em fotografias de arquivo, tendo em conta aspetos técnicos, éticos, organizacionais e tecnológicos. Em particular, foi dada prioridade à escolha de trabalhos que ajudassem a entender melhor como é que a descrição manual funciona, qual é o papel da indexação e da recuperação da informação, como é que os procedimentos de organização, normalização e interoperabilidade funcionam, e que questões éticas surgem com a representatividade, a marginalização e os enviesamentos históricos.

Por outro lado, foram excluídos os artigos que não se centrassem especificamente em fotografias

de arquivo ou que abordassem domínios adjacentes sem relação direta com a descrição arquivística.

Isto inclui estudos com foco em conjuntos de dados não relacionados, reconstrução 3D, realidade virtual, modelagem geoespacial, análise puramente algorítmica ou tecnológica, abordagens voltadas exclusivamente para património contemporâneo ou estudos visuais sem aplicação prática em metadados de arquivos fotográficos.

### 3.1.5 Análise das fontes

A análise das fontes foi realizada de forma sistemática, com a ajuda do gestor de referências *Zotero*[3] para organizar todas as referências e registar informação preliminar e também do *NotebookLM*[4] para explorar os artigos, ao nível do objetivo, método e relevância temática. Foram aplicados os critérios de inclusão e exclusão com base nos critérios definidos na subseção anterior.

## 3.2 PRISMA: *Query 1* e *Query 2*

Depois de executar todas as etapas do processo de pesquisa para a revisão sistemática de literatura (RSL), foi elaborado o diagrama Prisma para as perguntas de investigação Q1 e Q2. Para demonstrar de forma mais clara todos os passos realizados, elaboraram-se os diagramas que apresentam todas as fases, desde a identificação dos estudos, passando pelo processo de triagem e, por fim, a representação total dos estudos a incluir em cada revisão.

Os diagramas PRISMA permitem visualizar de forma clara o número de artigos excluídos em cada fase do processo de seleção, quer por duplicação, quer por não cumprirem os critérios de inclusão e exclusão definidos. Desta forma, torna-se possível identificar rapidamente todas as decisões tomadas ao longo do processo de revisão, para a revisão de Metadados em fotografia de arquivo, ver Figura 3.1, quer para a revisão de Inteligência Artificial: automatização de metadados, ver Figura 3.2.

[3] https://www.zotero.org/
[4] https://notebooklm.google/

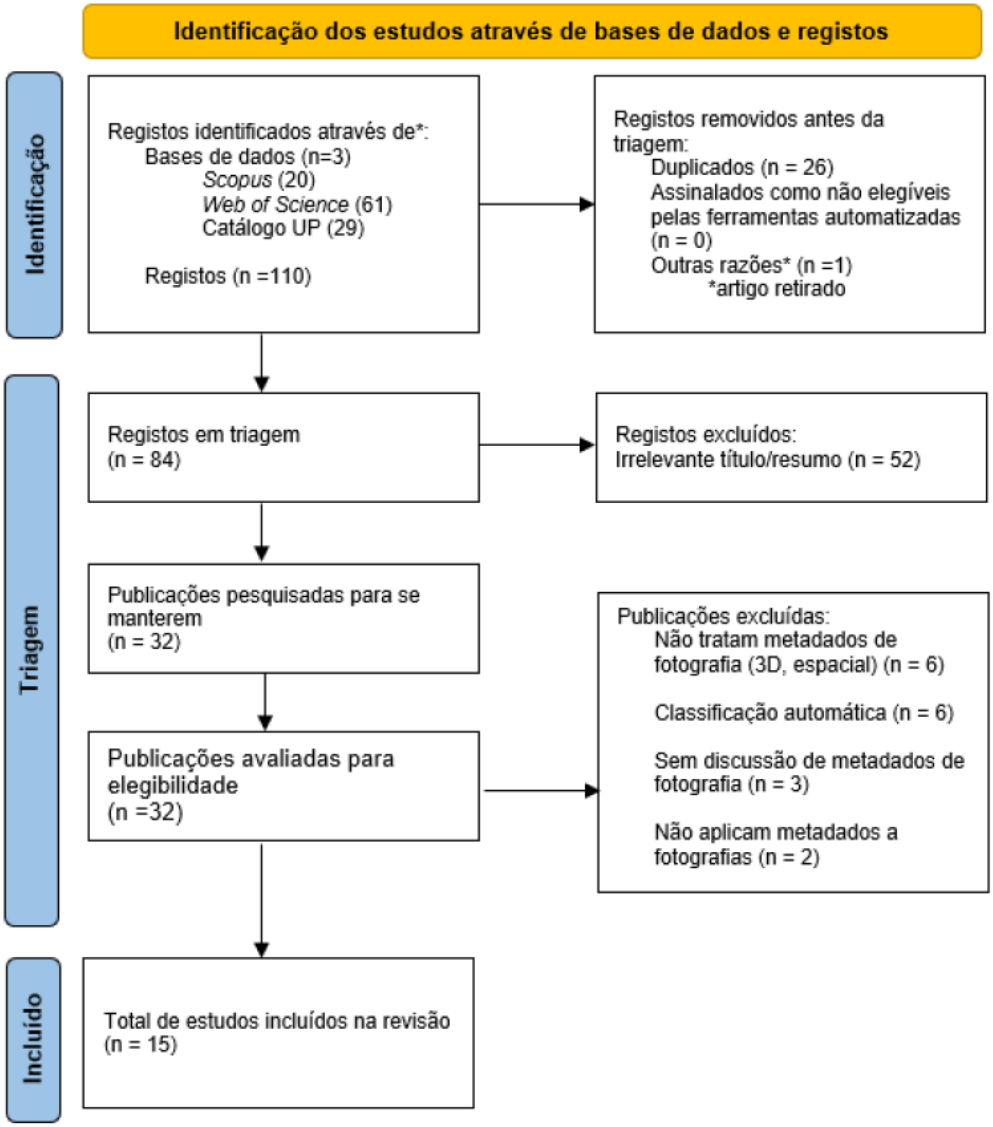



Figura 3.1: Diagrama PRISMA da *Query* 1

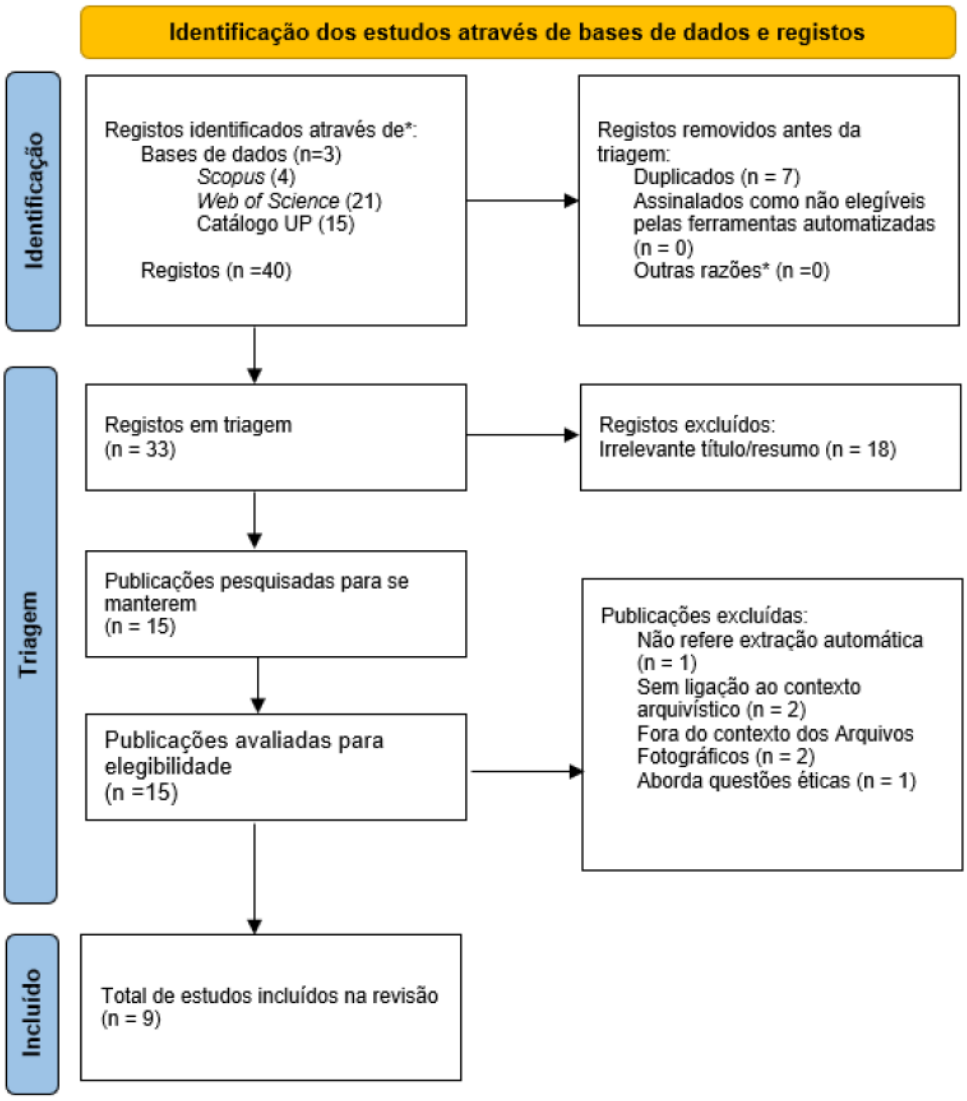



Figura 3.2: Diagrama PRISMA da *Query* 2

## 3.3 Investigação-ação

A investigação adotou uma metodologia de investigação-ação, aplicada em contexto institucional real, uma vez que envolveu a participação ativa do investigador na implementação, teste e avaliação dos processos metodológicos. A abordagem permitiu articular teoria e prática, aspeto fundamental em Ciência da Informação quando se pretende testar métodos e ferramentas em ambientes reais (Pickard, 2013).

No que toca ao ponto de vista epistemológico, o estudo está inserido num paradigma pragmático, centrado na resolução de problemas e na utilidade do conhecimento produzido. Este enquadramento metodológico possibilitou a compreensão de fenómenos complexos no seu contexto, ao mesmo tempo que permitiu a adaptação de procedimentos com base na evidência empírica recolhida (Saccol, 2009). O trabalho foi desenvolvido no Arquivo do Município de Viana do Castelo e respeitou as regras e procedimentos do arquivo, através de um estudo de caso, focado em analisar de

forma detalhada o caso específico com uma estratégia adequada para analisar fenómenos inseridos no contexto real, em que as fronteiras entre fenómeno e contexto não são claras (Yin, 2018).

A investigação-ação permitiu desenvolver o estudo em contexto de arquivo real, articulando intervenção prática, observação dos resultados e ajustes dos procedimentos em tempo real, esta abordagem mostrou-se adequada, uma vez que o percurso metodológico foi sendo adaptado em função das evidências recolhidas durante os testes, em particular após se verificar que o Módulo de Descrição de Imagem não incorporava de forma automática a lista de termos introduzidos pelo utilizador, nem gerava novas etiquetas para além das que já conhecia. Deste modo a metodologia manteve-se orientada para a resolução de um problema concreto: avaliar de que modo a Inteligência Artificial pode apoiar a descrição, a normalização e a recuperação da informação em retratos de estúdio.

### 3.3.1 Contexto institucional: Arquivo Municipal de Viana do Castelo

O estudo decorreu no Município de Viana do Castelo, na Unidade Orgânica de 3º Grau - Arquivo e Memória, que é o serviço municipal responsável pela preservação, descrição, gestão e difusão do acervo documental do município, constituindo o campo empírico da investigação e o estudo de caso selecionado, uma vez que foi identificada uma necessidade real relacionada com a descrição e recuperação da informação em fotografias, particularmente nos retratos de estúdio da Coleção Foto Roriz. A investigação assumiu uma abordagem de investigação-ação, dado que parte de um problema real do serviço e envolve uma intervenção prática no processo de descrição, através da seleção da amostra, construção de vocabulário controlado, testes com ferramentas de inteligência artificial, validação humana e avaliação dos resultados. Desta forma, o Arquivo Municipal de Viana do Castelo é entendido não apenas como contexto institucional, mas também como espaço de experimentação e observação sobre o contributo da inteligência artificial para a descrição de imagens de arquivo.

### 3.3.2 Caracterização da Coleção Foto Roriz

Em 2016, o acervo fotográfico do estúdio Foto Roriz foi doado ao Município de Viana do Castelo, passando a integrar o património documental do Arquivo. Esta coleção teve origem na atividade do estúdio fotográfico fundado em 1911 na Rua Gago Coutinho, em Viana do Castelo. Composto por milhares de imagens produzidas ao longo de mais de 100 anos, o acervo representa um testemunho visual único da história da cidade e da região.

A escolha da Coleção Foto Roriz para análise justifica-se pela presença significativa de retratos de estúdio, nos quais se encontram representadas pessoas, casais, famílias, crianças, grupos, mas também peças de vestuário, acessórios, cenários, entre outros objetos associados à fotografia comercial, tornando estas imagens uma fonte privilegiada para analisar formas de representação social e tradições culturais locais, ao mesmo tempo que oferecem um campo adequado para testar o contributo da inteligência artificial na descrição de fotografias, na identificação de elementos visuais e na recuperação da informação.

### 3.3.3 Caracterização e seleção da amostra

A amostra é constituída por 350 retratos de estúdio onde estão representadas pessoas, casais, famílias, crianças e grupos. Para além das pessoas representadas, as imagens incluem elementos visuais pertinentes para a descrição de fotografias, como vestuário, acessórios, poses, mobiliário, fundos de estúdio e outros objetos utilizados na composição do retrato.

Esta combinação entre uma tipologia comum e uma diversidade de elementos visuais torna a amostra adequada para avaliar o contributo da IA na descrição de fotografias e permite, em particular, testar a identificação automática de pessoas, objetos e atributos visuais, bem como analisar a necessidade de normalização terminológica e de validação humana no processo de descrição.

# Capítulo 4

# Conceção e Desenvolvimento do Modelo de Descrição

## 4.1 Preparação da amostra e levantamento

Seguido da seleção da amostra (descrita no capítulo anterior), deu-se início à fase de análise exploratória das imagens e preparação para a descrição arquivística. Esta fase teve como objetivo organizar os ficheiros digitais e enquadrar as fotografias no quadro de classificação do sistema de informação, mas também identificar os elementos visuais presentes nas imagens, para a construção do vocabulário controlado.

As fotografias selecionadas já se encontravam em formato digital, gravadas em CD, uma digitalização elaborada pelo Foto Roriz, em 2015 a partir do negativo original. Esta primeira fase consistiu na organização dos ficheiros numa tabela em Excel, onde se identificou de forma detalhada os diferentes pontos de acesso de forma exaustiva relativos a pessoas/agentes e objetos, como mostra a Imagem 4.1.

| | | | | | Pontos de acesso | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| código referência | CD | Título | Descrição geral | Imagem | Agente | Atividade | Objeto | Material | Cota | Condições de acesso | Condições de reprodução | Notas | Arquivista | Data da Descrição |
| 1 | CD1_1 | mulher adulta | Retrato de estúdio realizado pelo Foto Roriz, representa uma mulher adulta não identificada, representada em enquadramento de meio corpo e apoiada numa cadeira. A retratada apresenta cabelo encaracolado pelos ombros e usa brincos, colar de pérolas, pulseira e anel. Veste camisa escura de manga comprida. A fotografia foi realizada em contexto comercial, com fundo neutro e iluminação de estúdio controlada. Recurso fotográfico digitalizado. |  | pessoa (não identificada)<br>individual<br>feminino<br>adulto<br>cabelo encaracolado<br>corte pelos ombros<br>meio corpo | fotografia comercial | brincos<br>colar de pérolas<br>pulseira<br>anel<br>blusa escura de manga comprida<br>cadeira<br>fundo neutro<br>Estúdio Foto Roriz (Viana do Castelo) | recurso digitalizado | 4.77.6.1-13 | Acesso livre e gratuito | Disponível sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 Internacional (CC BY-NC-ND 4.0). É obrigatória a atribuição da autoria e do código de referência arquivística | | | 2026 |

Imagem 4.1: Grelha de análise exploratória dos retratos

Paralelamente deu-se início ao registo da amostra no sistema de informação, no fundo Foto Roriz, Sub-fundo Joaquim José Martins Carneiro, que seguiu a lógica de descrição multinível prevista pela ISAD(G). Deste modo procurou-se enquadrar cada fotografia numa estrutura hierárquica, partindo de um nível mais geral para o particular, para manter a ligação entre a fotografia e o respetivo contexto de produção, custódia e descrição.

Os retratos não foram tratados como simples ficheiros, mas como documentos com contexto, ou seja, documentos que fazem parte de uma coleção.

Foi atribuído o código de referência PT/MVCT-AMVCT/AFT/FR-1/04/A/1, com o objetivo de

identificar o retrato e localizá-lo no quadro de classificação do sistema de informação do arquivo. Assim, a sigla PT corresponde ao código do país, MVCT-AMVCT identifica a entidade detentora, e AFT/FR-1/04/A/1 corresponde ao código de referência local, definido de acordo com a organização interna da coleção.

Com o objetivo de facilitar a compreensão da estrutura do código atribuído, apresenta-se a sua decomposição na Tabela 4.1.

Tabela 4.1: Código de referência atribuído à amostra com base no ponto 3.1.1 da ISAD(G)

| **Composição** | **Siglas** | **Descrição** |
|---|---|---|
| Código do país | PT | Portugal. |
| Código da entidade detentora | MVCT-AMVCT | MVCT – Município de Viana do Castelo;<br>AMVCT – Arquivo Municipal de Viana do Castelo. |
| Código de referência local | AFT/FR-1/04/A/1 | AFT – Grupo de fundos Arquivos Fotográficos;<br>FR – Fundo Foto Roriz;<br>1 – Subfundo Joaquim Roriz Martins Carneiro;<br>04 – Coleção de negativos digitalizados;<br>A – IA sem intervenção;<br>1 – Retrato de estúdio na Foto Roriz: mulher adulta (post. 1911). |

Esta fase de preparação foi importante para garantir que a análise das imagens mantinha a sua ligação ao contexto. O retrato foi, assim, considerado em simultâneo como imagem digital, documento e testemunho visual integrado numa coleção.

Seguido desta organização, foram observados todos os retratos, um a um, onde a análise incidiu sobre todos os elementos visíveis com valor descritivo, nomeadamente, pessoas, vestuário, acessórios, cabelo, jóias, objetos, mobiliário e demais elementos do cenário de estúdio.

No que respeita às pessoas representadas, foram registadas sempre que possível e apenas com base em elementos visíveis, pessoas do género feminino, masculino, crianças, adultos, idosos. Observando também a posição e a composição do retrato, diferenciando situações como retrato individual, casal, grupo, pessoas em pé ou sentadas.

Também a análise ao vestuário e aos acessórios permitiu identificar elementos como vestidos, fatos, camisas, véus, sapatos, objetos religiosos, colares, brincos, fundos de cenário, cortinas, mesas, tapetes, entre outros elementos.

A referida observação foi representada numa grelha em Excel, que funcionou como instrumento intermédio entre a observação livre dos retratos e a posterior construção do vocabulário controlado. Os campos definidos permitiram registar todos os elementos visuais observados. Na Imagem 4.2 é possível verificar como se realizou o levantamento de todos os elementos visuais presentes em cada um dos retratos.

A análise realizada permitiu reconhecer padrões nas imagens e definir quais os termos que precisavam de normalização terminológica. Foi a partir deste levantamento que foi possível organizar categorias que serviram de base para a construção do vocabulário controlado e para as orientações do guia de descrição.

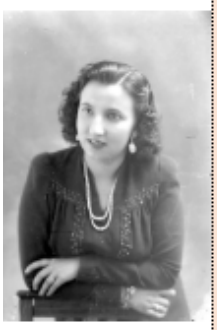
| | Pontos de acesso | | |
|---|---|---|---|
| Imagem | **Agente** | **Atividade** | **Objeto** |
|  | pessoa (não identificada)<br>individual<br>feminino<br>adulto<br>cabelo encaracolado<br>corte pelos ombros<br>meio corpo | fotografia<br>comercial | brincos<br>colar de pérolas<br>pulseira<br>anel<br>blusa escura de manga comprida<br>cadeira<br>fundo neutro<br>Estúdio Foto Roriz (Viana do Castelo) |
|  | pessoa (não identificada)<br>grupo 2<br>feminino<br>masculino<br>casal<br>adulto<br>cabelo curto<br>cabelo apanhado<br>corpo inteiro<br>em pé | fotografia<br>comercial | fato<br>casaco de trespasse 6x2<br>camisa branca<br>gravata com riscas<br>lenço lapela<br>sapato raso<br>sapato raso com laço<br>blusa escura com gola rendada branca<br>saia abaixo do joelho<br>fio<br>medalha (pendente)<br>custódia (pendente)<br>brincos<br>fundo pintado<br>carpete<br>Estúdio Foto Roriz (Viana do Castelo) |
|  | pessoa (não identificada)<br>individual<br>feminino<br>adulto<br>cabelo apanhado<br>corpo inteiro<br>em pé | fotografia<br>comercial | brincos<br>fio de contas<br>fio<br>medalha<br>blusa escura com gola rendada<br>avental de veludo com bolso e renda<br>saia abaixo do joelho<br>meias<br>chinelas lisas<br>pedestal de pé torneado<br>napron de croché<br>jarra com motivos<br>flor<br>fundo pintado<br>carpete<br>Estúdio Foto Roriz (Viana do Castelo) |

Imagem 4.2: Grelha de análise, campos utilizados na recolha de elementos visuais

## 4.2 Construção do vocabulário controlado

Esta primeira etapa resultou do levantamento dos elementos identificados na fase anterior, tendo como ponto de partida a análise individual de cada retrato e respetivo registo na grelha de análise, onde foram registados termos utilizados para descrever pessoas, vestuário, acessórios, cabelo, pose, composição, objetos e cenários de estúdio que permitiram construir uma base empírica para a criação de um vocabulário adaptado aos retratos de estúdio do acervo.

Inicialmente, os termos apresentavam uma expressão muito descritiva e composta, pois procuravam refletir com maior detalhe o que era observado nos retratos como, por exemplo, "blusa escura de manga comprida", "casaco de trespasse 6x2", "blusa escura com gola rendada branca", "jarra com motivos", "pedestal de pé torneado", "sapato raso com laço" ou "camisa branca".

Embora estas expressões fossem úteis para uma leitura inicial dos retratos, levantavam sérios problemas de consistência. O passo seguinte consistiu numa fase de revisão e normalização do conjunto de termos recolhido, na qual se eliminaram os duplicados, sinónimos, expressões longas e descritivas, e nos casos em que coexistiam termos equivalentes, foi selecionada uma forma preferencial, com o propósito de reduzir a dispersão terminológica.

Concluída a etapa de normalização, avançou-se para a etiquetagem no módulo de descrição de imagem do sistema de informação onde se constatou que seria mais adequado utilizar termos simples e diretamente associáveis a elementos visuais individualizáveis, como "menina", "homem", "mulher", "sapato", "camisola", "sobretudo", "cortina" ou "brinco". Na Figura 4.1 podemos ver o processo de evolução do termo descritivo ao termo simples, e na Tabela 4.2 alguns exemplos da conversão de termos compostos em termos simples.

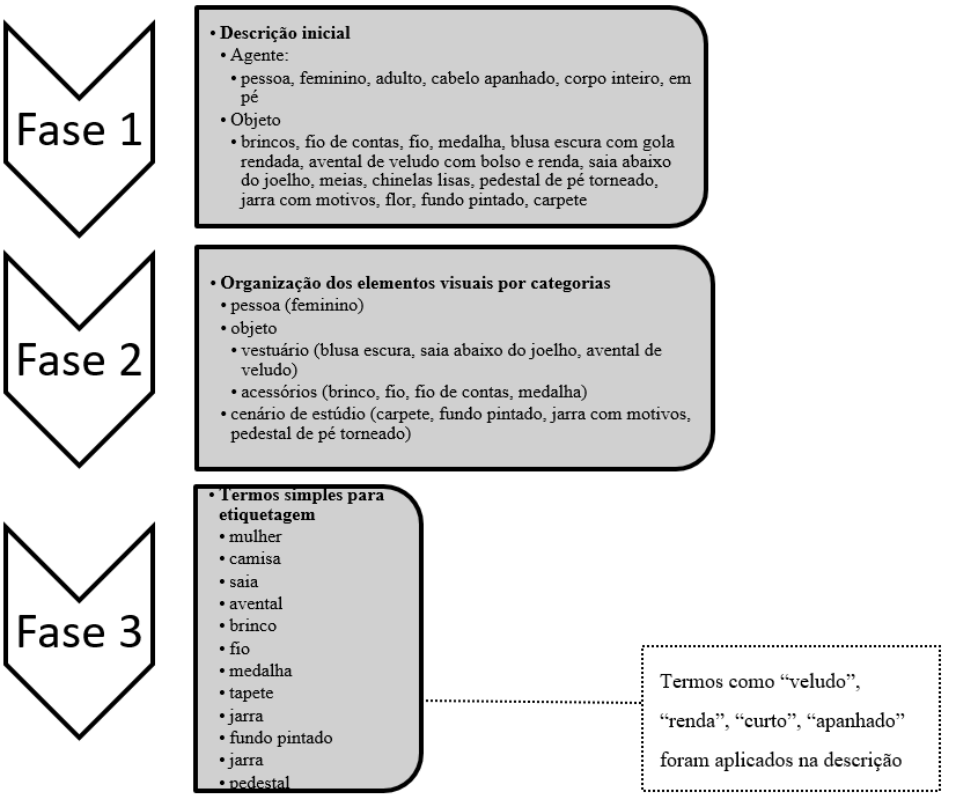



Figura 4.1: Processo de evolução do termo descritivo ao termo simples

Tabela 4.2: Exemplo de conversão de termos compostos em termos simples

| **Termo inicial** | **Termo normalizado (simples) para etiquetagem** | **Informação do atributo deslocada para o campo descrição geral** |
|---|---|---|
| "cabelo curto" | "cabelo" | "apresenta-se de cabelo curto" |
| "saia abaixo do joelho" | "saia" | "saia abaixo do joelho escura" |
| "sapato com atacador" | "sapato" | "sapatos com atacador escuros" |
| "sapato escuro" | "sapato" | "sapatos com atacador escuros" |
| "camisa clara" | "camisa" | "camisa clara" |

Assim, a lógica do vocabulário passou a privilegiar termos simples e atómicos, e os atributos como "claro", "escuro", "apanhado", "coberto", deixaram de integrar a etiqueta principal e passaram a ser registados no campo descrição geral, assim em vez de se criar uma etiqueta como "sobretudo claro" ou "vestuário (sobretudo)", privilegiou-se a etiqueta "sobretudo" sendo que o atributo "claro" registado como informação complementar, como se pode verificar nas Imagens 4.3, 4.4 e 4.5.

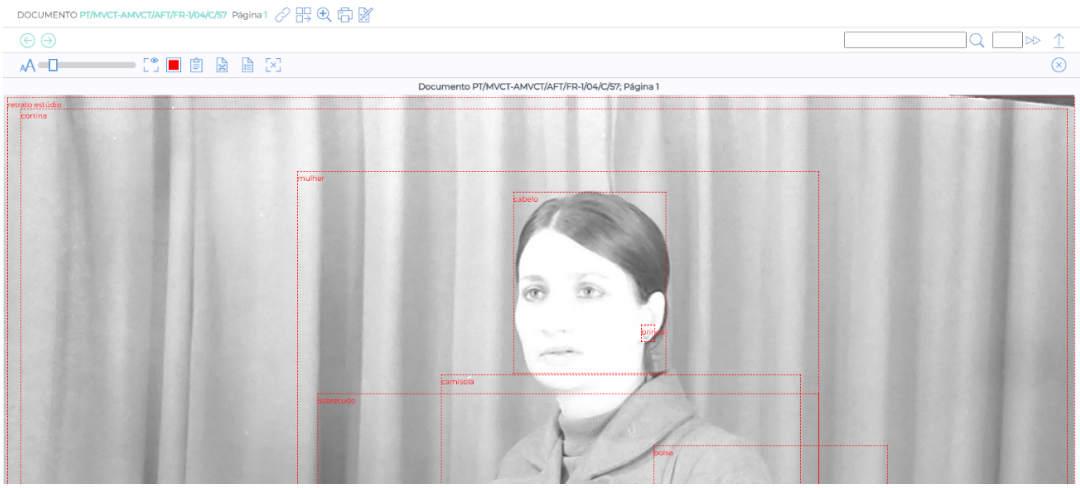



Imagem 4.3: Exemplo de etiquetagem em caixas delimitadoras




Imagem 4.4: Exemplo de etiqueta e caixa correspondente

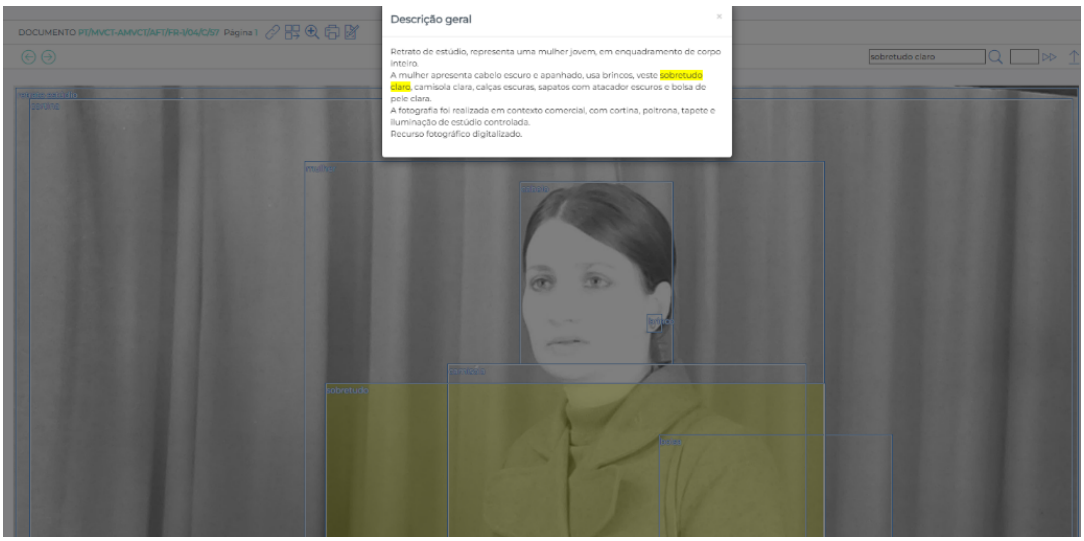



Imagem 4.5: Exemplo de descrição geral onde é mencionado o atributo “claro”

Esta foi uma decisão importante para adequar o vocabulário ao funcionamento do módulo e reforçar a coerência da descrição, pois a utilização de termos simples facilitou a etiquetagem dos

elementos visuais, reduzindo a ambiguidade o que permitiu uma maior uniformização na aplicação das etiquetas, enquanto a separação entre o termo principal e o atributo preservaram a possibilidade de uma descrição mais rica, sem comprometer a normalização.

Na medida do possível procurou-se ainda alinhar os termos com a lógica dos vocabulários controlados e dos tesauros utilizados em património cultural, nomeadamente com *Art & Architecture Thesaurus do Getty Research Institute*, que permitiu privilegiar conceitos mais genéricos evitando termos demasiado específicos, embora os elementos associados à tradição como "lenço de namorados" ou "traje de varina" fossem mantidos, mas sujeitos a maior cuidado descritivo. Na Tabela 4.3 podemos encontrar alguns exemplos da escolha de termos e do seu alinhamento com os termos do *AAT Getty*.

Tabela 4.3: Alinhamento de normalização com o *AAT Getty*

| **Termo** | **Termo do *AAT Getty*** | **Descrição do termo** | **URI** |
|---|---|---|---|
| "sobretudo" | sobretudo (vestuário) | Casacos usados sobre um terno ou outra peça de roupa principal, especialmente indicados para climas frios. | http://vocab.getty.edu/page/aat/300209839 |
| "camisa" | camisa (peça de vestuário) | Geralmente se refere a vários tipos de peças de vestuário principais para a parte superior do corpo, com ou sem mangas, usadas sobre ou por dentro da cintura de uma saia, calça ou similar. Normalmente possuem gola, que pode ser removível, frequentemente uma abertura frontal e, às vezes, bolsos. Especificamente, refere-se a uma peça de vestuário masculino com mangas compridas que terminam em punhos ou faixas nos pulsos. Originalmente, referia-se a roupas íntimas para a parte superior do corpo, feitas de linho, chita, flanela, seda ou outro material lavável; originalmente sempre usadas em contato direto com a pele. | http://vocab.getty.edu/page/aat/300212499 |
| "bolsa" | bolsa (acessório de vestuário) | Bolsas pequenas, médias ou grandes, macias ou rígidas, destinadas a transportar objetos pessoais e geralmente usadas como acessórios de vestuário feminino. Uma bolsa é fechada em todos os lados, exceto na parte superior, onde geralmente também pode ser fechada, e normalmente possui alças ou tiras para carregá-la no ombro ou na mão. | http://vocab.getty.edu/page/aat/300198926 |

Importa salientar que a lista de termos alcançada não deve ser entendida como uma lista fechada, mas como um ponto de partida desenvolvido a partir de uma amostra, com o intuito de servir de apoio à consistência da descrição de retratos de estúdio. A lista final poderá ser consultada no Anexo A.

## 4.3 Orientações para o guia de descrição

Uma vez concluída a versão final da lista de vocabulário controlado, sentiu-se a necessidade de criar um guia para aplicar na descrição de retratos de estúdio, uma vez que a lista por si só não era suficiente para garantir descrições consistentes, considerando que continuavam a surgir dúvidas acerca da escolha dos termos, grau de detalhe, distinção entre o que era elemento principal e atributo

e ou identificação de relações familiares ou sociais. Assim, pretende-se que estas orientações sirvam como um manual de boas práticas para garantir que todos descrevem os retratos de forma coerente, objetiva e recuperável, sem perder ligação ao contexto.

A construção do guia partiu dos resultados conseguidos no levantamento dos elementos visuais da amostra e da normalização dos termos realizada na fase anterior. Foram tidos em conta os princípios da descrição arquivística, o respeito pela proveniência, contexto, a ligação entre os diferentes níveis de descrição, e ainda articulada com a estrutura utilizada pela ISAD(G) e pelas Orientações para a Descrição Arquivística[1] (ODA).

O vocabulário foi construído seguindo boas práticas internacionais de descrição de património cultural, aproximando-se de referências como o Getty AAT, mas adaptado à realidade específica da coleção da amostra.

Para estas orientações também foram contempladas as limitações técnicas do Módulo de Descrição de Imagem, uma vez que a etiquetagem está organizada por *bounding boxes* a elementos específicos, definiu-se como principio orientador a utilização de termos simples e atómicos para a etiqueta principal, ficando os atributos, designadamente cor, padrão, material, para incluir no campo complementar descrição geral, estabelecendo assim três níveis de informação: o termo principal, que identifica o objeto/elemento visual "vestido", o atributo, que o qualifica "claro", "comprido" e a descrição geral que integra a informação no contexto da imagem "vestido de noiva comprido claro".

Outro aspeto considerado nestas orientações foi o tratamento face à incerteza, uma vez que a composição visual dos retratos de estúdio pode sugerir relações familiares, sociais, cerimoniais ou culturais sem que seja possível a sua confirmação através de informação documental complementar. Por isso o guia recomenda a utilização de descrições objetivas como "mulher e menina " em alternativa a "mãe e filha" sempre que não se consiga confirmar a relação familiar.

Também as descrições geradas por IA foram incluídas nas orientações, onde a verificação incide se a ferramenta identificou exatamente o que está na imagem, se a descrição está completa ou faltam elementos importantes, se foram utilizados sempre os mesmos termos para descrever os mesmos elementos, se os termos escolhidos são precisos e detalhados, se respeita as orientações e o contexto e se a descrição permite que se encontre com facilidade o retrato.

O guia de apoio à descrição de retratos de estúdio encontra-se no Anexo B.

## 4.4 Módulo de Descrição de Imagem

Finalizadas a fase de preparação da amostra, a construção do vocabulário controlado e as orientações para o guia de descrição, avançou-se para o teste do Módulo de Descrição de Imagem que recentemente integrou o sistema de informação do Arquivo.

Nesta nova etapa o objetivo foi avaliar como o Módulo de Descrição de Imagem poderia apoiar a descrição em retratos de estúdio, através de descrições automáticas de etiquetas, com aplicação de termos normalizados e posterior recuperação da informação. O teste incidiu na articulação entre a ferramenta, o vocabulário controlado e as orientações de descrição definidas segundo o guia, disponível no Anexo B.

[1] https://antt.dglab.gov.pt/wp-content/uploads/sites/17/2008/09/oda1_2_3.pdf

Numa fase inicial, o teste foi pensado como um processo faseado onde se pretendia observar o comportamento do módulo antes e depois da introdução da lista de termos, sendo planeado diferentes momentos, numa primeira versão, que designamos Versão A, foi solicitado ao módulo para gerar descrição em imagem de etiquetas para as 350 imagens da amostra. Na Versão B foram etiquetadas manualmente 44 imagens que não pertenciam à amostra, recorrendo à lista de termos construída na secção 4.2. Na Versão C, foi acionada novamente a função gerar descrição em imagem sobre as 350 imagens, para verificar se a etiquetagem manual dos termos tinha produzido alguma alteração nos resultados do módulo.

Durante este processo, verificou-se que a Versão C não apresentou alterações em relação à Versão A, o módulo continuou a recorrer às etiquetas já existentes no sistema, sem utilizar a lista de termos construída no estudo para gerar automaticamente etiquetas novas, e por esse motivo foi necessário ajustar o procedimento, em vez de se avançar para uma nova fase, procedeu-se à etiquetagem manual normalizada de 75 imagens da amostra, com o objetivo de criar uma base controlada para testar a recuperação da informação.

Deste modo, o teste ficou organizado em três versões principais. A Versão A correspondeu à geração automática inicial sobre as 350 imagens da amostra, a Versão B consistiu na etiquetagem manual de 44 imagens exteriores à amostra, utilizando a lista de termos e a Versão C integrou dois momentos: uma nova geração automática sobre as 350 imagens e, à posteriori a etiquetagem manual com base na lista de termos de 75 imagens da amostra. Na Tabela 4.4 podemos encontrar uma síntese das diferentes versões na fase de teste.

Tabela 4.4: Fase de teste do Módulo de Descrição de Imagem

| **Versão** | **Imagens** | **Procedimento realizado** | **Finalidade** |
|---|---|---|---|
| Versão A | 350 imagens da amostra | Aplicar a função “gerar descrição em imagem” do MDI | Analisar o comportamento do MDI face à amostra. |
| Versão B | 44 imagens fora da amostra | Etiquetagem manual da lista de termos | Fornecer ao módulo a lista de termos construída na Secção 4.2. |
| Versão C | 350 imagens da amostra | Aplicar a função “gerar descrição em imagem” do MDI | Verificar se os termos fornecidos se refletiam na amostra em análise. |
| | 75 imagens da amostra | Etiquetagem manual com base na lista de termos | Criar um conjunto controlado para teste de recuperação da informação. |

Em cada uma das fases, foram registados os dados necessários para acompanhar o procedimento: código de referência, versão do teste, descrição gerada pelo módulo, etiquetas automáticas e etiquetas criadas ou corrigidas manualmente, e ainda considerado o histórico de alterações do próprio módulo, que permite acompanhar as versões das etiquetas e da descrição geral associadas a cada imagem.

O módulo permite aceitar e rejeitar etiquetas, criar novas etiquetas, corrigir termos inadequados e inserir ou reescrever uma descrição geral da imagem. Todas as operações ficam registadas no sistema, o que permitiu distinguir a informação produzida automaticamente da informação resultante da intervenção humana.

Esta análise não tinha como propósito medir o desempenho computacional da ferramenta, mas compreender a sua utilidade no contexto dos arquivos.

No Anexo C foram incluídos *prints* das versões A, B e C, assim como os *prints* do histórico de alterações, etiquetas geradas automaticamente, criadas ou corrigidas manualmente e um exemplo de descrição geral.

## 4.5 Ajuste metodológico

Esta secção surge da necessidade de ajustar a metodologia após a fase de testes, correspondente ao início da Versão C, do Módulo de Descrição de Imagem. Esta necessidade decorreu da análise do funcionamento da ferramenta e da necessidade de adequar o percurso da investigação às limitações identificadas durante a realização dos testes.

Numa fase inicial, previa-se que a introdução manual de termos normalizados pudesse influenciar as etiquetas geradas automaticamente pelo módulo, mas na fase de testes verificou-se que o sistema continuava a recorrer às etiquetas já existentes, ignorando a lista de termos construída no estudo com a finalidade de gerar novas etiquetas.

Para contornar esta situação, o módulo passou a ser considerado sobretudo como um instrumento de apoio à etiquetagem manual, validação e recuperação da informação, e não como uma ferramenta de aprendizagem progressiva de vocabulário controlado. É importante ressalvar que este ajuste não alterou o objetivo principal da investigação, pelo contrário, permitiu esclarecer o papel do módulo de descrição de imagem no estudo e justificar a introdução de um teste complementar com o ChatGPT Vision.

## 4.6 Teste complementar ChatGPT Vision

Após os testes no módulo de descrição de imagem, sentiu-se a necessidade de ajustar a metodologia definindo-se um teste complementar com recurso à ferramenta ChatGPT Vision. Este teste teve como objetivo analisar de que forma uma ferramenta de IA generativa com capacidade de análise visual seria capaz de apoiar na descrição de retratos de estúdio, principalmente se orientada por humanos e com acesso a uma lista de termos.

Com este teste complementar não se pretende substituir o módulo de descrição de imagem, nem avaliar o ChatGPT Vision, mas apenas encontrar um ponto de comparação entre as duas formas de gerar descrições automáticas, primeiramente sem intervenção quer humana, quer terminológica e posteriormente com orientação humana, vocabulário controlado e critérios de descrição previamente definidos.

Foram selecionadas as três primeiras imagens da amostra para a realização deste teste. O número reduzido de imagens analisadas deveu-se às limitações de acesso da versão utilizada do ChatGPT Vision no momento do teste e, por isso, esta etapa é apenas um procedimento complementar e

exploratório. Ainda assim, as imagens incluídas representam diferentes situações visuais com retratos individuais ou em casal, vestuário, jóias e elementos de cenário.

O teste foi organizado em dois momentos, em que no primeiro cada imagem foi submetida à ferramenta apenas com a instrução:

> *"Descreva os elementos visíveis nesta fotografia"*.

No segundo, foi dada uma nova instrução à ferramenta:

> *"Agora descreva cada uma das fotografias utilizando o seguinte vocabulário controlado: homem, mulher, pedestal, jarra, arranjo de flores, brinco, colar de pérolas, fio, pendente, medalha, cadeira, camisa escura, casal grupo, sapato escuro, saia, fato, casaco, calça, lenço lapela, gravata, gola, cabelo apanhado, cabelo curto, fundo pintado, tapete, avental, chinela, retrato de estúdio, identifique apenas elementos visíveis, distinga homem, mulher, acessório de roupa, acessório de joalharia, cenário, pose"*.

Ao comparar os dois momentos foi possível aferir em que medida a introdução de orientações humanas e de vocabulário contribuía para descrições mais específicas, consistentes e adequadas ao contexto dos arquivos, uma vez que esta avaliação incidiu na identificação de elementos visíveis e na aproximação à lista de termos.

As evidências desta experiência, nomeadamente os *prompts* utilizados e as respostas geradas, podem ser consultadas no Anexo D.

# Capítulo 5

# Resultados e Discussão

O presente capítulo vai mostrar e analisar os resultados obtidos nas diferentes fases de investigação. Para isso, comparou-se a descrição humana normalizada com os resultados produzidos pelo módulo de descrição de imagem e pelo teste complementar com o ChatGPT Vision.

Com esta análise procurou-se perceber como cada abordagem identificou e descreveu os elementos presentes nos retratos de estúdio, tendo em conta a precisão dos termos utilizados, a consistência da descrição e os cuidados perante interpretações não comprovadas, a necessidade de validação humana e a utilidade das etiquetas para a recuperação da informação.

## 5.1 Resultados da descrição humana normalizada

Para este processo de análise foi considerada a amostra de 75 imagens com descrição humana e termos normalizados, os retratos etiquetados manualmente um a um, e as orientações definidas no guia de descrição. Foi esta etapa que permitiu criar uma base controlada para comparar os resultados produzidos pelo Módulo de Descrição de Imagem e pelo teste complementar com ChatGPT Vision.

A análise das 75 imagens mostrou que os termos mais recorrentes se relacionam sobretudo com a identificação das pessoas representadas, o vestuário e os acessórios, elementos que assumem particular importância nos retratos de estúdio, uma vez que contribuem para caracterizar a composição da imagem e funcionam como pontos de acesso relevantes para a recuperação da informação. Na Tabela 5.1 apresentam-se exemplos dos termos aplicados por categoria.

Tabela 5.1: Categorias e exemplos de termos aplicados

| **Categoria** | **Termos** |
|---|---|
| Pessoa | “mulher”, “homem”, “menina”, “menino”, “noiva”, “noivo” |
| Vestuário | “vestido”, “camisa”, “casaco”, “calça”, “fato”, “colete”, “saia” |
| Joalharia | “brinco”, “colar”, “fio”, “pulseira”, “anel”, “aliança” |
| Cenário | “cortina”, “tapete”, “mesa”, “revista”, “jarra”, “arranjo de flores”, “cadeira” |
| Acessórios | “gravata”, “lenço de lapela”, “véu”, “grinalda”, “bandolete” |

A organização por categorias permitiu separar os termos de acordo com a sua função na

descrição da imagem, desta forma foi possível distinguir termos utilizados para identificar pessoas, como "mulher" ou "homem", termos utilizados para descrever vestuário, como "vestido" ou "saia", termos relativos a acessórios como "gravata", termos de joalharia como "brinco" ou "anel", e termos associados ao cenário como "cortina", ou "tapete". Esta estrutura foi importante para evitar a dispersão terminológica e garantir que os termos aplicados funcionassem como pontos de acesso pesquisáveis.

De seguida analisou-se a frequência de alguns termos mais relevantes nas 75 imagens descritas manualmente, podendo os resultados ser consultados na Tabela 5.2.

Tabela 5.2: Frequência de termos na descrição humana

| **Termo** | **N.º de imagens** | **Percentagem da amostra** | **Imagens relevantes recuperadas** | **Precisão** | **Revocação** |
|---|---|---|---|---|---|
| mulher | 54 | 72% | 54 | 100% | 100% |
| homem | 25 | 33% | 25 | 100% | 100% |
| noiva | 10 | 13% | 10 | 100% | 100% |
| noivo | 8 | 11% | 8 | 100% | 100% |
| vestido | 24 | 32% | 24 | 100% | 100% |
| menina | 14 | 19% | 14 | 100% | 100% |
| menino | 10 | 13% | 10 | 100% | 100% |
| anel | 17 | 23% | 17 | 100% | 100% |
| brinco | 57 | 76% | 57 | 100% | 100% |
| brinco rainha | 11 | 15% | 11 | 100% | 100% |

Os dados mostram que os termos "brinco" e "mulher" foram os mais frequentes, surgindo, respetivamente, em 76% e 72% das imagens analisadas. Estes resultados evidenciam a relevância dos elementos associados à representação da pessoa e aos acessórios na descrição deste conjunto de retratos de estúdio.

É importante referir que os termos não são mutuamente exclusivos e a mesma imagem pode conter vários termos, pertencentes a categorias diferentes ou a níveis distintos de descrição, como por exemplo, uma imagem identificada com o termo "noiva" pode também ser descrita com os termos "mulher", "vestido", "brinco" ou "anel". Esta opção permite assegurar uma recuperação mais ampla, através de termos gerais e uma recuperação mais específica, através de termos associados a contextos ou objetos concretos.

Também o registo de uma descrição geral de cada imagem, permitiu completar os termos principais com atributos visuais, como cor, dimensão, posição, material, padrão ou grau de certeza, assim, a etiqueta "vestido" podia ser complementada, na descrição geral, com indicações como "claro", "comprido" ou "rendado" e a etiqueta "avental" podia ter o atributo "escuro" ou "veludo".

Esta articulação entre etiqueta e descrição geral aumentou a granularidade da descrição e as etiquetas funcionaram como pontos de acesso controlados, enquanto a descrição geral permitiu registar informação visual mais rica, útil para a interpretação da imagem e para a recuperação da informação.

Esta descrição permitiu reduzir a variação terminológica e aplicar os termos de forma consistente. Ao mesmo tempo, evidenciou a importância da validação humana, sobretudo na distinção entre elementos visíveis, atributos e interpretações. Sempre que a imagem não permitia confirmar uma relação familiar, social ou cerimonial, privilegiaram-se descrições cautelosas.

Embora útil, esta abordagem também apresentou limitações, pois a descrição manual exige tempo, depende da visibilidade dos elementos na imagem e pode envolver decisões interpretativas, mesmo quando apoiada em orientações e vocabulário controlado. Além disso, a aplicação de etiquetas simples não representa, por si só, relações complexas entre pessoa, objeto e atributo.

Ainda assim, esta etapa foi essencial para responder aos objetivos do estudo, uma vez que permitiu construir uma base normalizada, testar a aplicação prática do vocabulário controlado e preparar a avaliação da recuperação da informação no sistema.

## 5.2 Resultados do módulo de descrição de imagem

Testar o módulo de descrição de imagem permitiu perceber como a ferramenta funciona na prática quando aplicada a retratos de estúdio em contexto real. A análise ocorreu sobre três aspetos principais: gerar etiquetas de forma automática, a possibilidade de criar, aceitar ou rejeitar etiquetas e criar descrições e a utilidade das etiquetas para recuperar imagens no sistema.

Na abordagem inicial, o módulo foi capaz de gerar etiquetas para as imagens analisadas, etiquetas essas que resultavam dos termos que o sistema já trazia e permitiram identificar alguns elementos visíveis nas fotografias, nomeadamente “pessoa”, “vaso”, “ planta envasada”, “gravata”, “sofá”. Contudo, também permitiu a intervenção humana e, desta forma, foi possível aceitar ou rejeitar etiquetas, acrescentar novas etiquetas e corrigir termos inadequados, como por exemplo o termo “telemóvel”, quando na imagem aparecia, na realidade, um “livro de orações”.

Este exemplo, demonstra como a validação humana é essencial para retificar inconsistências ou erros de descrição, quando os elementos têm aspetos visuais semelhantes e que podem induzir ao erro.

Além dessas funcionalidades, o módulo de descrição de imagem também permite acompanhar as diferentes alterações efetuadas ao longo do registo, através do histórico, sendo possível acompanhar o percurso de cada imagem e perceber que informação tinha sido gerada automaticamente e o que tinha sido acrescentado ou corrigido manualmente. Este aspeto revelou-se importante, pois mostrou que o módulo não funciona apenas como ferramenta automática, mas também como espaço de revisão, correção e validação da informação produzida.

No entanto, foi identificada uma limitação relevante, nomeadamente quando o módulo ignorou os termos fornecidos na Versão B, mesmo depois de terem sido aplicadas etiquetas controladas a 44 imagens exteriores à amostra. Como já mencionado, ao aplicar novamente a função “gerar descrição de imagem” na Versão C sobre as 350 imagens, o módulo voltou a produzir os mesmos resultados da Versão A, o que indica que o módulo recorre às etiquetas já existentes no sistema e não aprendeu progressivamente a partir da lista de termos construída no estudo.

- Resultados obtidos

Na Versão A, o módulo gerou automaticamente etiquetas para as 350 imagens da amostra,

nas quais os resultados permitiram identificar alguns elementos visuais, mas, em muitos casos, as etiquetas eram genéricas, como “pessoa” e “gravata”, isto é, generalidades que se revelaram insuficientes para a descrição de retratos de estúdio, sobretudo quando era necessário distinguir vestuário, acessórios, joalharia, pose, composição ou elementos de cenário. Na Imagem 5.1 apresenta-se um exemplo de etiquetagem automática.




Imagem 5.1: Resultado da etiquetagem automática (Versão A)

Na Versão B, foram etiquetadas manualmente 44 imagens fora da amostra, recorrendo à lista de termos, que nesta fase foi possível fornecer ao módulo, no entanto, embora a ferramenta se tenha mostrado funcional para etiquetagem de elementos visuais, essa ação não se refletiu no comportamento ao gerar etiquetas na fase seguinte. Na Imagem 5.2 apresenta-se um exemplo da etiquetagem na fase de treino.




Imagem 5.2: Resultado da etiquetagem da amostra de treino (Versão B)

Na Versão C, foi acionada novamente a função gerar descrição automática sobre as 350 imagens da amostra, tendo os resultados sido semelhantes aos da Versão A, o que confirmou que o módulo não incorporou a lista de termos fornecida pelo utilizador, como se observa na Imagem 5.3.

Considerando esta limitação, avançou-se para a Versão C1, na qual foram etiquetadas manual-

mente 75 imagens da amostra com termos normalizados.




Imagem 5.3: Resultado da etiquetagem na Versão C, para a referência AFT/FR-1/04/C/23

Nesta etapa, foram etiquetadas manualmente 75 imagens da amostra com termos normalizados. Esta etapa permitiu criar um conjunto controlado de imagens etiquetadas e descritas de acordo com as orientações do guia, tornando possível testar a recuperação da informação a partir dessas etiquetas e descrições, ver Imagem 5.4.




Imagem 5.4: Resultado da etiquetagem manual na Versão C1

Os resultados indicam que o módulo é útil quando utilizado com intervenção humana. A ferramenta pode apoiar a descrição, a correção, a validação e a pesquisa, mas a qualidade dos resultados depende da etiquetagem manual de termos normalizados e da revisão técnica do arquivista.

- Limitações identificadas

O facto do módulo não incorporar automaticamente a lista de termos fornecida pelo utilizador foi identificado como a principal limitação, pois embora a lista construída no estudo tenha servido para apoiar a etiquetagem manual, ela não foi utilizada pelo sistema para gerar novas etiquetas de forma automática.

Também a generalidade das etiquetas produzidas foi entendida como limitação, pois em muitos casos, os termos gerados eram demasiado amplos para descrever com rigor os retratos de estúdio, por exemplo, uma etiqueta genérica pode indicar a presença de uma pessoa ou de um objeto, mas não permite representar com detalhe o vestuário, os acessórios, os atributos ou a composição da imagem.

Outro aspeto verificado foi a ausência de relações entre elementos, podendo existir uma etiqueta para “mulher”, outra para “saia” e outra para “casaco”, mas o sistema não indica que a saia e

o casaco pertencem à mulher representada, esta é uma limitação particularmente relevante em imagens com várias pessoas, porque dificulta uma recuperação mais precisa da informação, como se verifica na Imagem 5.5.




Imagem 5.5: Resultado de pesquisa para "mulher saia casaco"

Por fim, os termos utilizados no módulo funcionam como etiquetas textuais internas, sem ligação a identificadores externos, como os do AAT Getty. Esta ausência limita a interoperabilidade semântica e a possibilidade de reutilização futura dos dados em ambientes mais estruturados.

Deste modo, o teste demonstrou que o Módulo de Descrição de Imagem pode ser uma ferramenta útil de apoio à descrição e à recuperação da informação, mas não substitui o trabalho humano e o seu contributo depende da existência de vocabulário controlado, de critérios claros de descrição e da validação técnica realizada pelo arquivista.

## 5.3 Resultados do teste com ChatGPT Vision

Este teste complementar permitiu comparar as descrições produzidas sem orientação terminológica e com instruções humanas e lista de termos, embora o teste tenha incidido apenas sobre três imagens da amostra, os resultados permitiram observar diferenças relevantes entre os dois momentos.

No primeiro momento, sem orientação, as descrições produzidas foram sobretudo narrativas e livres, a ferramenta identificou elementos visíveis nas fotografias, como pessoas, vestuário, pose, acessórios e cenário, mas fê-lo sem uma estrutura estável e sem recurso a vocabulário controlado. Em alguns casos, a descrição aproximou-se de uma leitura interpretativa da imagem, sugerindo relações, contextos ou intenções que não são comprovados apenas pelos elementos visuais, como se pode verificar na imagem do Anexo D, onde se refere *"sugerindo uma pose de casal ou de familiares próximos"*.

No segundo momento, já com orientação humana, as descrições tornaram-se mais organizadas

e próximas das categorias definidas no guia de descrição, após as indicações a ferramenta passou a reconhecer de forma mais clara. Exemplo disso é que em vez de "pessoa" já identificou "mulher". No entanto, apenas numa lógica de hierarquia, como mostra a Imagem 5.6 e como se pode verificar no Anexo D: vestuário, acessórios, joalharia, pose, composição e cenário.

Verificou-se também maior cuidado na formulação de hipóteses, com recurso a expressões como "possível", "provável" ou "aparente" quando a identificação não era segura.

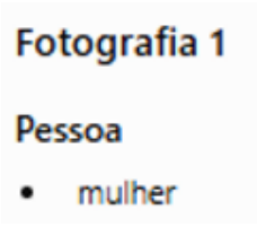



Imagem 5.6: Resultado do teste com ChatGPT Vision

Ao comparar os dois momentos, verificou-se que a orientação humana teve impacto positivo na qualidade das descrições, ou seja, as instruções permitiram reduzir a dispersão terminológica, aumentar a especificidade dos elementos identificados e aproximar a descrição dos critérios definidos para os retratos de estúdio. Ainda assim, os resultados continuaram a exigir validação humana, sobretudo para confirmar termos, corrigir eventuais interpretações e adequar a descrição ao contexto do arquivo.

No âmbito desta análise o ChatGPT Vision mostrou-se útil como ferramenta complementar de apoio à observação visual e a gerar etiquetas de descrição, apesar de na curta amostra a sua aplicação mostrar-se mais consistente quando orientada por instruções claras, vocabulário controlado e critérios de cuidado interpretativo.

Os *prompts* utilizados e as respostas geradas encontram-se no Anexo D.

## 5.4 Comparação entre abordagens

Além da qualidade descritiva, da precisão e da recuperação da informação, foi também considerado o esforço humano associado a cada abordagem, conforme apresentado na Tabela 5.3.

A Tabela 5.3 evidenciou que as abordagens automáticas exigiram um esforço humano baixo inicialmente, mas apresentam limitações na consistência, na precisão e na contextualização. Em contrapartida, as abordagens manuais ou orientadas por vocabulário exigem maior investimento humano, mas produzem descrições mais consistentes, exaustivas e adequadas à recuperação da informação. Então a comparação entre abordagens mostra que a IA pode apoiar e acelerar o processo descritivo, mas a qualidade arquivística depende de estratégias humanas, nomeadamente normalização e validação.

Para avaliar o impacto da intervenção humana na exaustividade da descrição, fez-se uma comparação da versão A, produzida automaticamente pelo sistema, com a versão C, posteriormente editada na amostra de 75 imagens, onde se evidenciou que na versão A foram atribuídas 277 etiquetas, correspondendo a uma média de 3,7 etiquetas por imagem, sendo predominantes termos genéricos como "s/descrição" e "pessoa", e após a edição manual, a Versão C passou a integrar

Tabela 5.3: Indicadores de esforço na descrição das imagens

| **Teste** | **N.º de imagens** | **Tipo de intervenção** | **Esforço humano** | **Observação** |
|---|---|---|---|---|
| MDI (Versão A) | 350 | Automática | Baixo | Recorreu aos termos já existentes no módulo |
| MDI (treino: fornecimento da lista de termos) | 44 | Manual | Elevado | Inserção da lista de termos, após a sua construção |
| MDI (Versão C) | 350 | Automática | Baixo | Resultados idênticos aos da Versão A |
| Etiquetagem manual da amostra | 75 | Manual | Elevado | Maior controlo |
| ChatGPT Vision (sem orientação) | 3 | Automática e assistida | Médio | Descrições geradas sem vocabulário controlado |
| ChatGPT Vision (com orientação) | 3 | Automática, assistida e validada | Médio/elevado | Orientação por vocabulário e validação humana |

1678 etiquetas, com uma média de 22,4 etiquetas por imagem, através da utilização de termos mais específicos e granulares, como “homem”, “mulher”, “menino”, “menina”, “vestido”, “casaco”, “saia”, “meia”, “sapato”, “fato”, “anel” e “brinco”.

Através da Tabela 5.4 foi possível demonstrar um aumento expressivo da exaustividade descritiva após a intervenção humana. Enquanto a Versão A apresentou uma descrição reduzida e pouco diferenciada, centrada em termos genéricos, a Versão C1 editada permitiu identificar diferentes elementos visuais e representar melhor pessoas, vestuário, acessórios e objetos presentes nas fotografias. Este aumento de granularidade reforçou a importância da validação e edição humana para melhorar a precisão terminológica, ampliar os pontos de acesso e favorecer a recuperação da informação.

Assim, a passagem de 277 para 1678 etiquetas evidencia que a intervenção humana não apenas aumentou a quantidade de informação registada, como também qualificou a descrição, tornando-a mais específica, estruturada e útil para a recuperação das fotografias.

Tabela 5.4: Comparação do número de etiquetas entre versões

| **Versão** | **Tipo de descrição** | **N.º de etiquetas** | **Média de etiquetas por imagem** | **Caracterização** |
|---|---|---|---|---|
| Versão A | Automática | 277 | 4 | Etiquetas genéricas, com predominância dos termos “pessoa” e “sem descrição”. |
| Versão C | Manual | 1 678 | 22 | Descrição com maior granularidade e utilização de termos específicos relativos a pessoas, vestuário, acessórios e joalharia. |

## 5.5 Impacto na recuperação da informação

De modo a conseguir avaliar o impacto da descrição na recuperação da informação, foram realizadas pesquisas com três diferentes níveis de filtro. A pesquisa 1 por um termo visual, como "fato", permitiu recuperar um conjunto mais amplo de imagens dentro da coleção Foto Roriz. Verificou-se que a introdução progressiva de novos filtros reduziu o número de resultados, tornando a pesquisa mais precisa e ajustada à intenção do utilizador.

Na Tabela 5.5 é possível evidenciar que a combinação progressiva de etiquetas permitiu refinar a pesquisa e aumentar a precisão da recuperação. Na pesquisa 1, realizada com a etiqueta "fato", foram recuperadas 20 imagens, correspondendo a um conjunto amplo de fotografias onde este elemento de vestuário estava presente. Ao aplicar outro filtro com a etiqueta "menino", o número de resultados reduziu-se para 3 imagens, restringindo a pesquisa aos retratos em que surgia simultaneamente menino vestido com fato, e, por fim, um novo filtro com a etiqueta "vela" permitiu recuperar apenas 1 imagem, correspondente ao registo que reunia os três elementos pesquisados: "fato", "menino" e "vela".

Tabela 5.5: Demonstração de recuperação de informação

| **Pesquisa** | **Filtro** | **Termo pesquisado** | **N.º imagens** | **Resultados** |
|---|---|---|---|---|
| Pesquisa 1 | Menção de responsabilidade + etiqueta | "Foto Roriz" + "fato" | 20 | Recuperou todas as imagens com a etiqueta "fato" dentro do produtor Foto Roriz |
| Pesquisa 2 | Menção de responsabilidade + etiqueta | "Foto Roriz" + "fato" + "menino" | 3 | Reduziu os resultados às imagens associadas a "menino" de "fato" |
| Pesquisa 3 | Menção de responsabilidade + etiqueta | "Foto Roriz" + "fato" + "menino" + "vela" | 1 | Recuperou o registo que reúne os pontos de acesso selecionados |

Assim, a redução progressiva dos resultados, de 20 para 3 e depois para 1 imagem, demonstra o impacto da granularidade descritiva na recuperação da informação, como se pode ver na Imagem 5.7. Demonstra também que a existência de etiquetas específicas permite ao utilizador partir de uma pesquisa mais geral e chegar a um resultado mais preciso, ajustado à sua intenção de pesquisa.

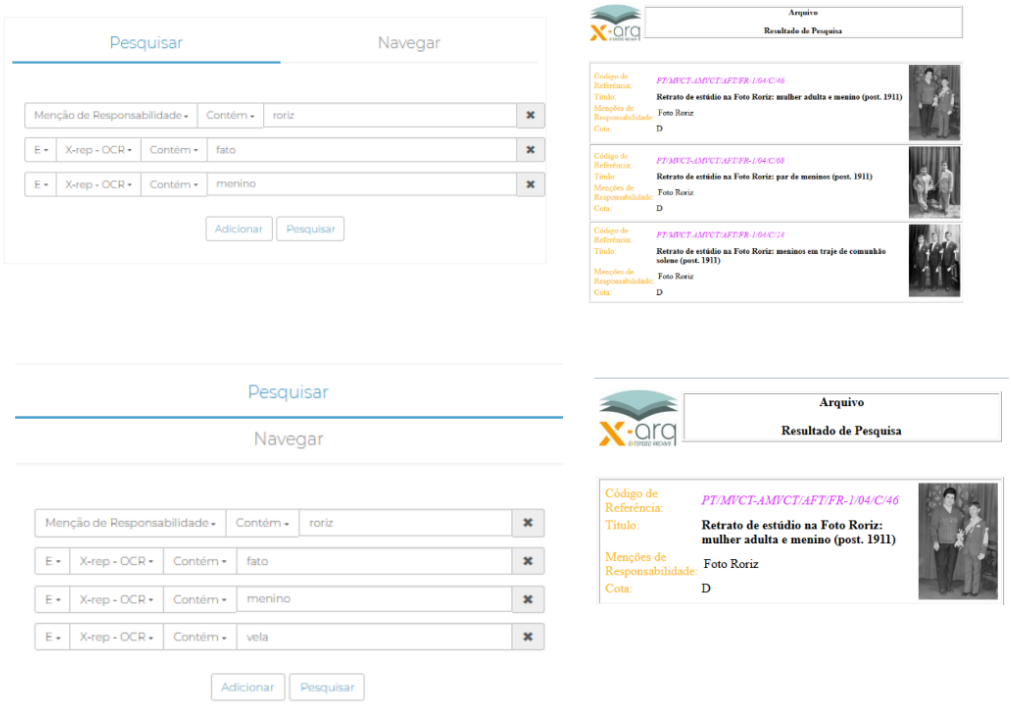



Imagem 5.7: Resultados do segundo e terceiro filtros de pesquisa

## 5.6 Discussão dos resultados

A Tabela 5.6 faz uma síntese dos resultados da investigação que demonstram que a IA pode apoiar a descrição de fotografias, sobretudo na identificação inicial de elementos visuais e na criação de novos pontos de acesso à informação, no entanto o teste realizado no MDI revelou limitações importantes: a descrição automática foi genérica na Versão A e na a Versão C não apresentou alterações evidentes, apesar da introdução prévia de termos normalizados através da amostra de treino. Este resultado indicou que o sistema não funcionou como um modelo com aprendizagem progressiva, mas como uma ferramenta de descrição geral.

Por sua vez, o teste complementar com o ChatGPT Vision permitiu observar uma diferença interessante entre a descrição sem orientação e a descrição orientada. Na primeira, os resultados foram mais livres e menos consistentes, na segunda, as respostas tornaram-se mais completas, mais próximas do vocabulário controlado e mais úteis para a recuperação da informação, resultado que confirma que a IA pode ser mais eficaz quando integrada num processo orientado por regras, vocabulários controlados e validação humana.

Mas, foi a descrição humana que se mostrou mais adequada para garantir coerência terminológica, contexto arquivístico e interpretação cultural da imagem, a sua maior exaustividade, quer na etiquetagem, quer na descrição geral, permitiu representar um número mais elevado de elementos visuais, como pessoas, vestuário, acessórios, pose, cenário e atributos, e essa diferença tornou-se evidente no aumento da média de etiquetas atribuídas, que passou de cerca de **4 etiquetas** por

Tabela 5.6: Síntese de resultados

| **Ferramenta/teste** | **Impacto** |
|---|---|
| MDI (Versão A > Versão C) | Gerou etiquetas de acordo com a sua lista interna e, após uma fase de treino, não se verificou aprendizagem progressiva |
| ChatGPT Vision (sem orientação) | Gerou descrições livres, muito narrativas e com grande variedade de termos |
| ChatGPT Vision (com orientação) | Adaptou-se às instruções fornecidas, melhorando a qualidade da descrição |
| Descrição humana | Processo exaustivo, quer na etiquetagem, quer na descrição geral |
| Etiquetagem por *bounding boxes* | Apenas permite uma etiqueta por caixa, impedindo a associação de atributos a objetos e contexto |

imagem na descrição automática para cerca de **22 etiquetas** por imagem na descrição humana. Foi essa riqueza descritiva que aumentou os pontos de acesso à imagem e, consequentemente, favoreceu a recuperação da informação. Por se tratar de uma seleção controlada e validada dos termos, esta abordagem contribuiu para melhorar a precisão dos resultados recuperados.

Para finalizar, a limitação de uma única etiqueta por caixa mostrou que a descrição visual baseada apenas em objetos não é suficiente para representar fotografias de forma completa. A imagem contém relações entre pessoas, vestuário, acessórios, cenário, pose e contexto social, por isso, abordagens como o CIDOC CRM e o *Linked Data* tornam-se relevantes como possibilidade futura, permitindo representar relações entre entidades, atributos e contextos.

## 5.7 Implicações para os arquivos

Após a interpretação dos resultados obtidos, a investigação demonstrou que a IA pode efetivamente apoiar na descrição de fotografias de arquivo, no entanto, a validação humana é um fator crucial, visto que a literatura reforça que a descrição de documentos de arquivo exige contexto, normalização de termos, coerência e validação, dimensões que continuam dependentes do conhecimento profissional do técnico de arquivo.

No caso analisado, a IA revelou utilidade na identificação inicial de elementos visuais, mas mostrou limitações ao nível da interpretação de atributos, relações, contexto cultural e normalização. Assim, constatou-se que a IA deve ser entendida como uma ferramenta auxiliar, sobretudo útil para acelerar tarefas repetitivas e apoiar a primeira camada de identificação visual, mas a integração destas ferramentas deve ser um processo controlado, pois a descrição apenas se tornará útil se for devidamente articulada com vocabulário controlado, regras de normalização e validação humana. Sem estes elementos, existe o risco de serem produzidas descrições pouco consistentes, ambíguas ou difíceis de recuperar. Pode, ainda, verificar-se uma implicação mais crítica, nomeadamente a reescrita involuntária da história, ao serem privilegiadas interpretações automáticas e descontextualizadas.

Também os sistemas de arquivo devem evoluir para modelos mais estruturados e interoperáveis e a simples atribuição de etiquetas isoladas é insuficiente para representar a complexidade das

fotografias, pelo que devem ser pensadas articulações futuras com identificadores persistentes, grafos de conhecimento e modelos de representação como o CIDOC CRM.

Em síntese, adotar IA nos arquivos só produzirá resultados consistentes e relevantes se for enquadrada por vocabulários controlados, regras de normalização, critérios de validação e supervisão humana, de modo a garantir que a tecnologia reforça, e não fragiliza os princípios arquivísticos de contexto, rigor, acesso e recuperação da informação.

# Capítulo 6

# Conclusões

A presente investigação teve como objetivo analisar o contributo da IA na descrição de fotografias, a partir da análise da Coleção Foto Roriz, que integra o acervo documental do AMVCT. A investigação partiu de uma amostra de retratos de estúdio, onde se procurou compreender de que forma a descrição automática de imagens poderia apoiar na identificação de elementos visuais e melhorar a recuperação da informação, quando em articulação com descrição, vocabulário controlado e validação humana.

Os resultados permitem concluir que a IA pode constituir um instrumento útil de apoio à descrição de fotografias, principalmente na identificação inicial de elementos visuais evidentes, como pessoas, objetos, vestuário, acessórios ou elementos de cenário, mas verificou-se que a sua aplicação apresenta limitações significativas quando não existe aprendizagem efetiva, integração com vocabulários controlados, reconhecimento de atributos ou capacidade de representar relações semânticas entre os elementos identificados.

A investigação confirmou as perspetivas apresentadas na revisão de literatura, na perspetiva de que a IA pode ampliar os pontos de acesso à informação e apoiar processos de descrição, mas substituir a mediação do profissional de informação, a descrição ou validação humana continua a ser essencial para garantir contexto, consistência terminológica, interpretação cultural. Esta articulação entre automatização e validação humana revelou-se particularmente importante nas fotografias com informação visual, documental, social e cultural.

Recorrer a vocabulário controlado demonstrou ser uma condição importante para melhorar a consistência da descrição e a recuperação da informação, tal como defendido na literatura sobre normalização, metadados e interoperabilidade. A simples etiquetagem automática de objetos não é suficiente para garantir descrições consistentes, no entanto, é necessário controlar os termos utilizados, distinguir conceitos, atributos e relações para assegurar que a descrição responde às necessidades de pesquisa dos utilizadores.

Também a pertinência da polirrepresentação na descrição de fotografias, a combinação entre descrição geral, termos controlados, etiquetas de objetos e validação humana permite representar a imagem de forma mais completa e aumentar as possibilidades de recuperação, abordagem que se mostrou mais adequada do que uma descrição exclusivamente automática ou exclusivamente textual.

Como principal contributo, a investigação apresenta que a IA deve ser entendida como uma ferramenta auxiliar no processo de descrição de imagens. O seu valor reside na possibilidade de apoiar a identificação preliminar de elementos visuais, acelerar tarefas repetitivas e sugerir pontos de acesso adicionais, no entanto, a qualidade dos resultados depende da existência de critérios de normalização, supervisão humana e enquadramento arquivístico.

No seguimento deste trabalho de investigação e num ponto de vista de trabalho futuro, considera-se pertinente aprofundar soluções que permitam uma maior integração semântica dos metadados fotográficos, isto é, uma forma de descrição que não se limite à atribuição de etiquetas isoladas, mas que permita identificar e relacionar os vários elementos associados às imagens. Para tal, a articulação com vocabulários controlados permitiria garantir o uso consistente dos termos, evitando variações como "brinco", "brincos" ou "acessório", enquanto os identificadores persistentes permitiriam associar cada termo, entidade ou objeto a uma referência estável e não ambígua, um termo como "brinco" ou "retrato de estúdio" poderia ser ligado a um conceito normalizado, por exemplo no *AAT Getty*, reforçando a interoperabilidade e a reutilização da informação.

Recorrer a *Linked Data* seria pertinente, pois permitiria ligar os dados produzidos no sistema de arquivo a outros conjuntos de dados disponíveis em rede, tornando as descrições mais abertas, pesquisáveis e reutilizáveis por diferentes sistemas. A ligação a grafos de conhecimento permitiria organizar esses dados como uma rede de relações, representando, por exemplo, que uma fotografia pertence a uma coleção, foi produzida por determinado estúdio, representa uma pessoa, inclui determinados objetos ou acessórios e está relacionada com um lugar ou tradição cultural.

Também recorrer a modelos como o CIDOC CRM assume particular relevância, uma vez que oferecem uma estrutura normalizada para representar informação patrimonial, relacionando documentos, pessoas, objetos, eventos, lugares e períodos cronológicos, e por isso a sua aplicação permitiria reforçar a contextualização arquivística das fotografias, melhorar a interoperabilidade entre sistemas e favorecer uma recuperação da informação mais precisa.

No final a investigação permitiu concluir que a IA pode contribuir para melhorar a descrição e a recuperação da informação em fotografias, desde que integrada num processo controlado, validado e semanticamente orientado. Concluiu-se, também, que o futuro da descrição das fotografias não passa pela substituição do trabalho humano, mas pela combinação crítica entre tecnologia, normalização terminológica, conhecimento arquivístico e validação profissional.

# Referências bibliográficas


Abgaz, Y., Souza, R., Methuku, J., Koch, G., & Dorn, A. (2021). A Methodology for Semantic Enrichment of Cultural Heritage Images Using Artificial Intelligence Technologies. JOURNAL OF IMAGING, 7(8). https://doi.org/10.3390/jimaging7080121

Alkemade, H., Claeyssens, S., Colavizza, G., Freire, N., Lehmann, J., Neudecker, C., Osti, G., & Van Strien, D. (2023). Datasheets for Digital Cultural Heritage Datasets. JOURNAL OF OPEN HUMANITIES DATA, 9. https://doi.org/10.5334/johd.124

Arquivo Municipal de Viana do Castelo. (n.d.). Arquivo Municipal de Viana do Castelo. Câmara Municipal de Viana do Castelo. https://arquivo.cm-viana-castelo.pt/

Bachi, V., Fresa, A., Taes, S., Truyen, F., Fink, E., Wallace, M., Patias, P., Doulamis, A., Ioannides, M., Brumana, R., & Martins, J. (2018). PHOTOCONSORTIUM: Opening up the Riches of Europe's Photographic Heritage. In LECT NOTES COMPUT SC (Vol. 11196, pp. 585–598). Springer International Publishing AG. https://doi.org/10.1007/978-3-030-01762-0_51

Bushey, J. (2024). Envisioning Archival Images with Artificial Intelligence. Archeion (Warszawa), 2024(125), 33–54. https://doi.org/10.4467/26581264ARC.24.007.20202

Câmara, A., de Almeida, A., Caçador, D., & Oliveira, J. (2023). Automated methods for image detection of cultural heritage: Overviews and perspectives. ARCHAEOLOGICAL PROSPECTION, 30(2), 153–169. https://doi.org/10.1002/arp.1883

Dalia Ortiz Pablo, Badri, S., Norén, E., & Nötzli, C. (2023). Bias mitigation techniques in image classification: Fair machine learning in human heritage collections. arXiv.Org.

Davidson, A., & Reid, P. (2022). Digital storytelling and participatory local heritage through the creation of an online moving image archive: A case-study of Fraserburgh on Film. JOURNAL OF DOCUMENTATION, 78(2), 389–415. https://doi.org/10.1108/JD-09-2020- 0157

De Jager-Loftus, D. P. (2024). Sharing South Dakota's cultural heritage: Harvesting digital collections into the Digital Public Library of America and beyond. Art Libraries Journal, 49(1), 31–36. https://doi.org/10.1017/alj.2023.32

Eiler, F., Graf, S., & Dorner, W. (2018). Artificial intelligence and the automatic classification of historical photographs. 852–856. Scopus. https://doi.org/10.1145/3284179.3284324

Fornaro, P., e Chiquet, V. (2020). Artificial intelligence for content and context metadata retrieval in photographs and image groups. In Proceedings of ISeT Archiving 2020 (pp. 79–82). Society for Imaging Science and Technology. https://doi.org/10.2352/issn.2168-3204.2020.1.0.79

Garliauskas, A., e Šoliūnas, A. (1998). Aprendizagem e reconhecimento de padrões visuais por sujeitos humanos e sistemas de inteligência artificial. Informatica: An International Journal of Computing and Informatics, 9(4), 415–424. https://doi.org/10.3233/INF-1998-9402

Gowri, S., Homavazir, M., Setia, N., e Gupta, M. (2025). SMART ARCHIVING SYSTEMS FOR PHOTOGRAPHIC COLLECTIONS. 6(1s), 84–94. Scopus.https://doi.org/10.29121/shodhkosh.v6.i1s.2025

Hansson, K., & Dahlgren, A. (2022). Crowdsourcing historical photographs: Autonomy and control at the Copenhagen City Archives. COMPUTER SUPPORTED COOPERATIVE WORK-THE JOURNAL OF COLLABORATIVE COMPUTING AND WORK PRACTICES, 31(1), 1–32. https://doi.org/10.1007/s10606-021-09418-z

Hrechaniuk, Y., & Sparrman, A. (2024). Ethics in research practice: Young people, pictures, and archives. Archives and Records, 45(3), 306–323. Scopus. https://doi.org/10.1080/23257962.2024.2394084

Late, E., Ruotsalainen, H., & Kumpulainen, S. (2024). Image searching in an open photograph archive: Search tactics and faced barriers in historical research. International Journal on Digital Libraries, 25(4), 715–728. Scopus. https://doi.org/10.1007/s00799-023-00390-1

Lee, B. (2021). Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset. Digital Humanities Quarterly, 15(4).

Lo Turco, M., Calvano, M., & Giovannini, E. (2019). DATA MODELING FOR MUSEUM COLLECTIONS (A. Cardaci, F. Fassi, & F. Remondino, Eds.; WOS:000466553500060; Vols. 42–2, Issue W9, pp. 433–440). https://doi.org/10.5194/isprsarchives-XLII-2-W9-433-2019

Odumosu, T. (2020). The Crying Child On Colonial Archives, Digitization, and Ethics of Care in the Cultural Commons. CURRENT ANTHROPOLOGY, 61, S289–S302. https://doi.org/10.1086/710062

Oruc, P. (2025). Cultural Heritage as AI Training Data. IIC-INTERNATIONAL REVIEW OF INTELLECTUAL PROPERTY AND COMPETITION LAW, 56(8), 1455–1459. https://doi.org/10.1007/s40319-025-01659-5

Park, J. (2023). “We want to know more than that”: Lessons learnt from the public workshop on collections data at the V &A. MUSEUM MANAGEMENT AND CURATORSHIP, 38(2), 141–156. https://doi.org/10.1080/09647775.2022.2111331

Pickard, A. J. (2013). Research methods in information. Facet Publishing.

Pierce, R. (2019). The Female Gaze? Postmodernism and the Search for Women in the Digitized Photographic Collections of Swedish Memory Institutions. Open Information Science, 3(1), 61–75. Scopus. https://doi.org/10.1515/opis-2019-0005

Saccol, A. Z. (2009). Um retorno ao básico: Compreendendo os paradigmas de pesquisa e sua aplicação na pesquisa em administração [Back to basics: Understanding the research paradigms and their application in management research]. Revista de Administração da UFSM, 2(2), 250–269. https://doi.org/10.5902/198346592671

Sarkar, M., & Biswas, S. (2020). Exploring archives space: An open source solution for digital archiving. DESIDOC Journal of Library and Information Technology, 40(5), 272–276. Scopus. https://doi.org/10.14429/djlit.40.05.16330

Schaffland, A., & Schöning, J. (2023). Faroe Islands rephotography image registration dataset. DATA IN BRIEF, 47. https://doi.org/10.1016/j.dib.2023.108979

Tian Xin, Li, N., Ai Nisha, Gao Songhua, & Chen, L. (2025). Intelligent Identification of Rural Productive Landscapes in Inner Mongolia. Computers (Basel), 14(12), 565. https://doi.org/10.3390/computers14120565

Triques, M., Roa-Martínez, S., & de Albuquerque, A. (2024). Cultural Heritage Images as Digital Objects from the Perspective of Polyrepresentation: An Europeana Image Classification Pilot Review. INVESTIGACION BIBLIOTECOLOGICA, 38(100), 29–48. https://doi.org/10.22201/iibi.24488321xe.2024.100.58866

Truyen, F., & Fresa, A. (2016). PHOTOCONSORTIUM: Digitizing Europe's photographic heritage. 10058 LNCS, 409–419. Scopus. https://doi.org/10.1007/978-3-319- 48496-9_33

VAN Wissen, L., Vriend, N., Nijssen, A., Vereecken, L., & DEN Engelse, M. (2025). FAIR Photos—Transforming a Collection of Two Million Historical Press Photos into Five Star Data. JOURNAL OF OPEN HUMANITIES DATA, 11. https://doi.org/10.5334/johd.271

Wolff, J., Probst, A., & Bodenschatz, E. (2024). Bringing researchers together with IIIF. Art Libraries Journal, 49(1), 37–43. https://doi.org/10.1017/alj.2023.33

Wu, J. (2024). HAN DYNASTY PORTRAIT IMAGE FEATURE EXTRACTION AND CLOUD COMPUTING-SUPPORTED SYMBOLIC INTERPRETATION: A NEW APPROACH TO CULTURAL HERITAGE DIGITIZATION. SCALABLE COMPUTING- PRACTICE AND EXPERIENCE, 25(6), 4804–4813. https://doi.org/10.12694/scpe.v25i6.3299

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). Sage Publications.

# ANEXOS

# Anexo A – Lista de termos da amostra

Tabela A.1: Lista de termos da amostra

| **Pessoas** | **Composição** | **Posição** | **Enquadramento** | **Vestuário** | **Acessórios de roupa** | **Joalharia** | **Objetos religiosos** | **Cenário** | **Atributos** |
|---|---|---|---|---|---|---|---|---|---|
| pessoa | casal | sentado | meio corpo | vestido | gravata | brinco | rosário | cadeira | claro |
| mulher | grupo | ajoelhado | corpo inteiro | vestido de noiva | lenço de lapela | brinco rainha | livro de orações | poltrona | escuro |
| homem | família | em pé | | vestido de primeira comunhão | luva | fio | crucifixo | cortina | comprido |
| menino | noivos | | | vestido de comunhão solene | bolsa | pulseira | genuflexório | tapete | curto |
| menina | | | | saia | | colar | | fundo pintado | padrão geométrico |
| bebé | | | | saia plissada | | colar de pérolas | | fundo neutro | padrão animal |
| indeterminado | | | | camisa | | anel | | mesa | riscas |
| idoso | | | | camisola | | aliança | | pedestal | bordado |
| idosa | | | | casaco | | pendente cruz | | revista | coberto |
| | | | | sobretudo | | pendente medalha | | jarra | richelieu |
| | | | | fato | | pendente custódia | | arranjo de flores | |
| | | | | colete | | | | brinquedo | |
| | | | | calça | | | | flor | |
| | | | | traje regional | | | | | |

# Anexo B – Guia para a descrição de imagens

**Guia prático de Descrição de Imagem**

## 1. Objetivo

Este guia tem como objetivo apoiar a descrição normalizada e a validação de fotografias, de modo a garantir maior normalização, melhor recuperação da informação e maior consistência entre descrição humana, etiquetagem manual e ferramentas de IA.

Aplica-se, em particular, ao Módulo de Descrição de Imagem e aos retratos de estúdio; no entanto, é aplicável a outras coleções de fotografia.

## 2. Orientações gerais

A descrição deve ser objetiva: apenas devem ser registados os elementos visíveis na imagem ou sustentados por evidência arquivística comprovada.

Devem evitar-se interpretações não verificáveis, como relações familiares, profissão ou estatuto social, a menos que exista uma fonte que as confirme.

Para além da descrição arquivística prévia, que deve seguir as orientações das normas ISAD(G), ISAAR(CPF) e ODA, a descrição de imagem deve articular dois níveis complementares:

| Campo | Função |
|---|---|
| Descrição geral | Registar a leitura narrativa da imagem, incluindo pessoas, pose, vestuário, acessórios, cenário, composição e atributos. |
| Etiquetas | Identificar elementos visuais através de termos simples e normalizados. |

## 3. Orientações para a descrição geral

Preferencialmente, o texto narrativo da descrição geral deve seguir uma ordem estável: pessoas representadas, enquadramento, descrição da pessoa ou das pessoas, seguindo-se o cenário, contexto de produção e informação técnica do recurso.

Exemplo:

> *"Retrato de estúdio, representa um casal adulto, em enquadramento de corpo inteiro.*
>
> *A mulher adulta apresenta cabelo apanhado e usa brincos rainha, fio com pendentes de medalha, alfinete, veste camisa de manga comprida clara, com gola rendada clara, avental com bordado richelieu escuro, saia com riscas abaixo do joelho escura, meias rendadas claras e chinelas de salto médio escuras.*
>
> *O homem adulto apresenta-se de cabelo curto e usa anel e fio com medalha na lapela, veste fato claro composto por calça e casaco, camisa clara, gravata com padrão geométrico e sapatos claros.*
>
> *A fotografia foi realizada em contexto comercial, com fundo pintado e iluminação de estúdio controlada.*
>
> *Recurso fotográfico digitalizado.*

## 4. Orientações para a etiquetagem

As etiquetas devem ser simples e normalizadas, correspondendo cada uma a um único conceito visual.

Devem privilegiar-se termos como: mulher, homem, saia, casaco, fato, cadeira e brinco.

Devem evitar-se etiquetas compostas, como: avental com bolsos, vestido claro, brinco grande ou banco estofado.

Apenas pessoas e objetos devem ser etiquetados.

Atributos como cor, material, forma ou padrão devem ser registados preferencialmente no campo Descrição Geral.

Os atributos devem igualmente usar termos atómicos e normalizados, como: claro, escuro, comprido, curto ou bordado para que o processo de recuperação seja bem-sucedido.

## 5. Procedimento para a normalização

O processo de normalização deve seguir três fases: deve iniciar com o registo livre de todos os elementos observados, depois ser organizado pelas diferentes categorias e por fim ser convertido em termos simples.

| **Descrição inicial** | **Categorias** | **Etiqueta normalizada** | **Descrição geral** |
|---|---|---|---|
| Avental escuro de veludo com bolso e renda | Vestuário: avental;<br>Material: veludo;<br>Atributos: renda, escuro | avental | Avental escuro de veludo com bolso e aplicação de renda |

O processo de etiquetagem deve organizar-se por categorias estáveis e sempre que possível estar alinhados com vocabulários controlados, como o *AAT Getty*[1] .

[1] https://www.getty.edu/research/tools/vocabularies/aat/

| **Categorias** | **Termos** |
|---|---|
| Pessoas | pessoa, mulher, homem, menino, menina, bebé, indeterminado |
| Composição | casal, noivos, família, grupo |
| Posição | sentado, em pé, ajoelhado |
| Enquadramento | meio corpo, corpo inteiro |
| Vestuário | vestido, saia, fato, camisa, camisola, colete |
| Acessórios de roupa | gravata, lenço de lapela, luva, bolsa |
| Joalharia | brinco, colar, pulseira, anel, pendente |
| Objetos religiosos | livro de orações, rosário, cruxifixo |
| Cenário | tapete, cortina, pedestal, mesa, brinquedo |

## 6. Validação de etiquetas

Após o processo de descrição automática da imagem, algumas etiquetas podem surgir previamente identificadas pelo sistema. Estas etiquetas devem ser sempre verificadas, uma vez que podem estar corretas, incompletas ou incorretamente atribuídas ao elemento visual representado.

A validação consiste em confirmar se a etiqueta sugerida corresponde efetivamente ao objeto ou elemento delimitado pela caixa na imagem. Para proceder à revisão, deve clicar-se sobre a caixa delimitadora associada à etiqueta. A partir dessa seleção, é possível editar o termo sugerido e substituí-lo por uma etiqueta normalizada, de acordo com o vocabulário controlado definido.

Quando a etiqueta estiver correta, deve ser mantida ou ajustada apenas para garantir a forma normalizada do termo. Quando a etiqueta estiver incorreta, ambígua ou não corresponder ao elemento assinalado, deve ser rejeitada. Nos casos em que o sistema não permite apagar diretamente a etiqueta, esta deve ser assinalada como rejeitada, de forma a não ser considerada na descrição validada.

## 7. Exemplo de aplicação

1. No final do processo de etiquetagem deverá aparecer as etiquetas editadas e validadas na coluna Etiquetas, como mostra na imagem B.1

2. Ao clicar em cima de uma etiqueta deverá aparecer na imagem um retângulo a selecionar o que representa, como mostra na imagem B.2.

3. No final do processo de etiquetagem, todas as etiquetas selecionadas deverão estar visíveis na imagem, como mostra a imagem B.3.

4. Ao fazer a pesquisa por um termo ou mais na caixa de pesquisa, este deverá aparecer selecionado na imagem, como mostra a imagem B.4, para uma pesquisa como os termos fato e claro.

5. A descrição geral deverá seguir a matriz fornecida, utilizando os termos da lista de termos. No final da descrição o resultado deverá ser como o que mostra na imagem B.5.

6. Todas as etiquetas devem ser validadas ou rejeitadas, na imagem B.6 mostra um termo

validado aceite e na imagem B.7 um termo validado rejeitado.

7. Todas as alterações efetuadas na descrição de imagem ficam registadas em diferentes versões, como se pode ver na imagem B.8.




Imagem B.1: Exemplo de aplicação das etiquetas




Imagem B.2: Exemplo de representação de etiqueta na imagem

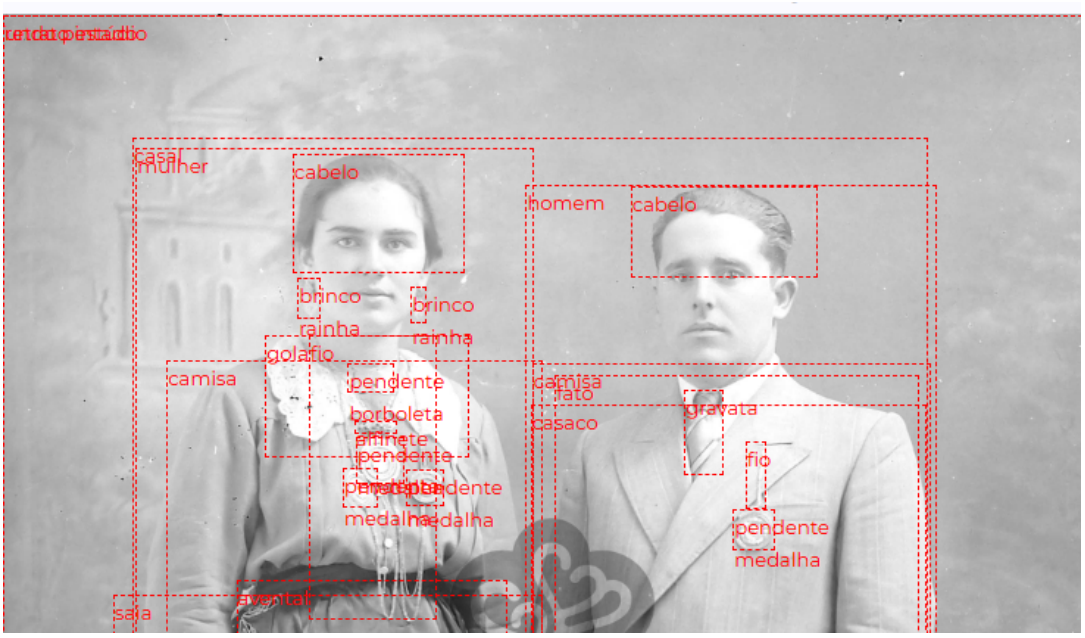



Imagem B.3: Resultado do processo de etiquetagem na imagem




Imagem B.4: Resultado da pesquisa na imagem

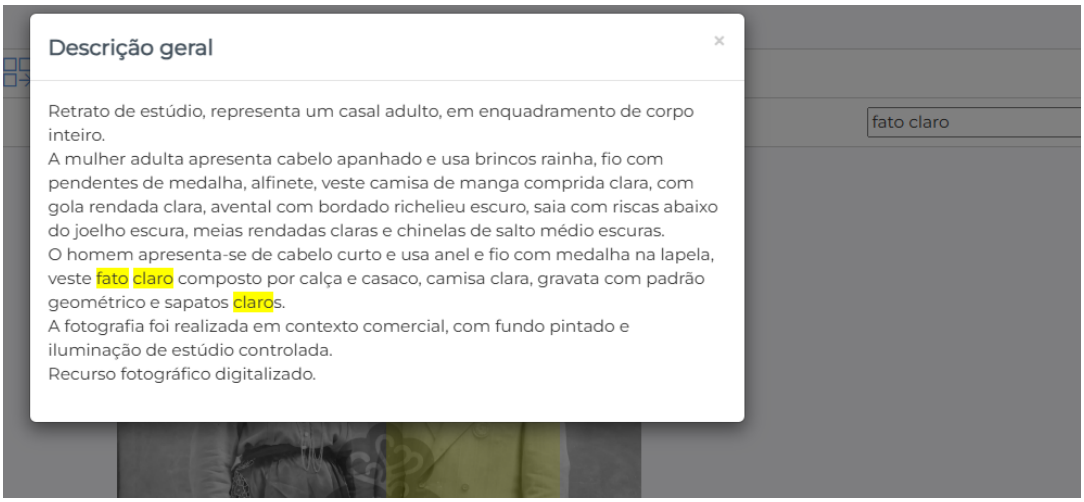



Imagem B.5: Resultado da pesquisa visível na descrição geral




Imagem B.6: Termo validado (**aceite**)

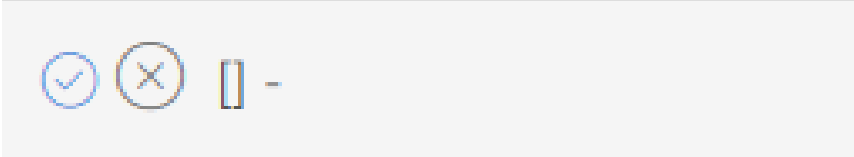



Imagem B.7: Termo validado (**rejeitado**)

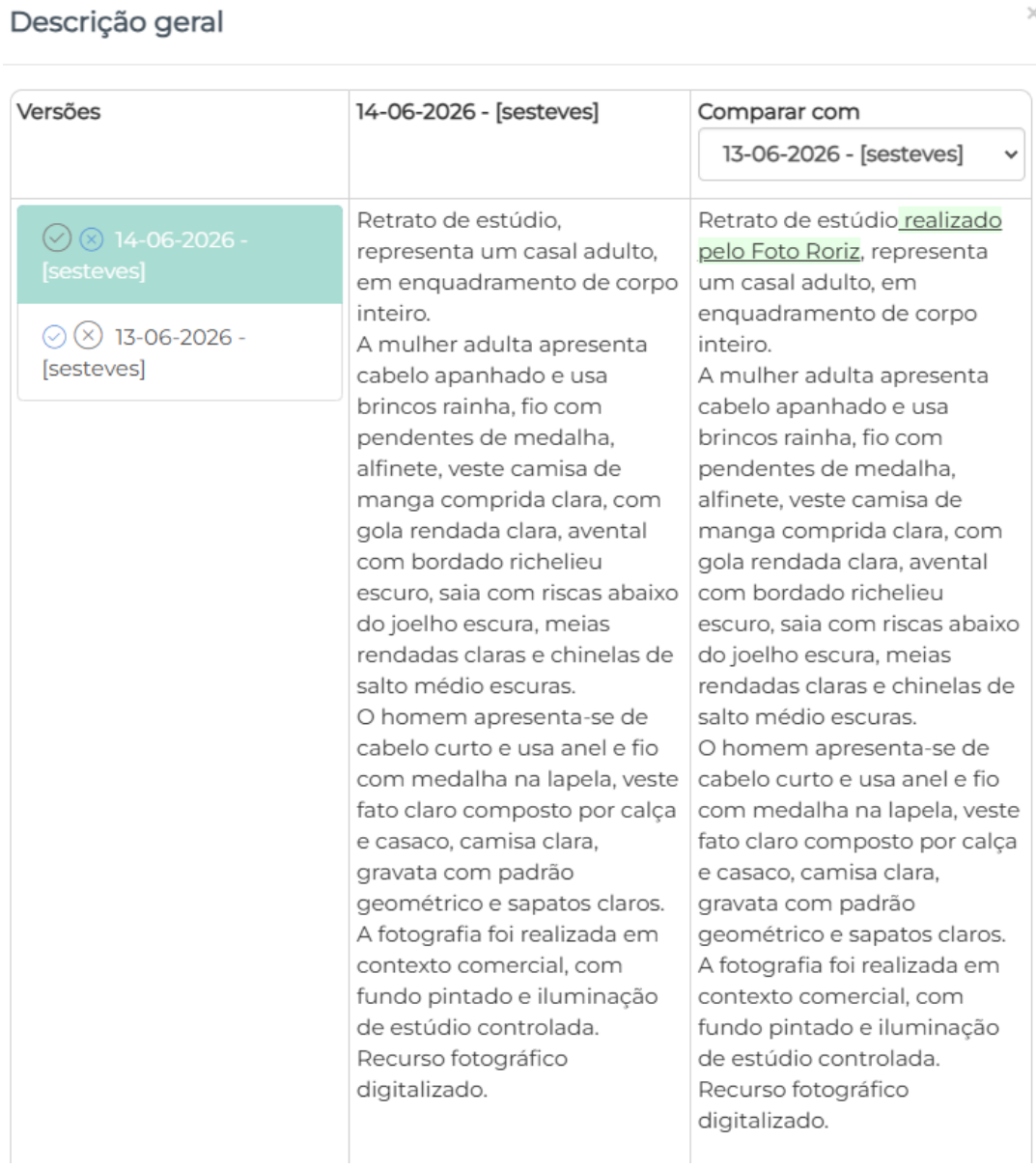



Imagem B.8: Registo do histórico de versões do campo descrição geral

# Anexo C – Resultados do Módulo de Descrição de Imagem

## 1. Versão A – Etiquetas geradas automaticamente




Imagem C.1: Exemplo das etiquetas geradas automaticamente na Versão A

## 2. Versão B - Etiquetas atribuídas pelo utilizador




Imagem C.2: Exemplo das etiquetas atribuidas pelo utilizador na Versão B

## 3. Versão C - Descrição de imagem no MDI




Imagem C.3: Descrição de imagem no MDI, Versão C1

## 4. Etiquetas geradas pelo MDI

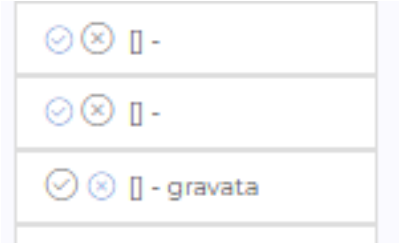



Imagem C.4: Etiquetas geradas pelo MDI

# 5. Etiquetas criadas pelo utilizador

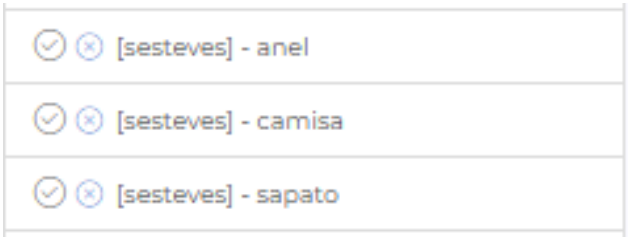



Imagem C.5: Etiquetas criadas pelo utilizador no MDI

# 6. Exemplo de descrição geral e histórico de versões




Imagem C.6: *Print* do campo descrição geral e histórico de versões

# Anexo D – Resultados do ChatGPT Vision

## Teste 1 - sem orientação

*Teste 1: Input (prompt):*

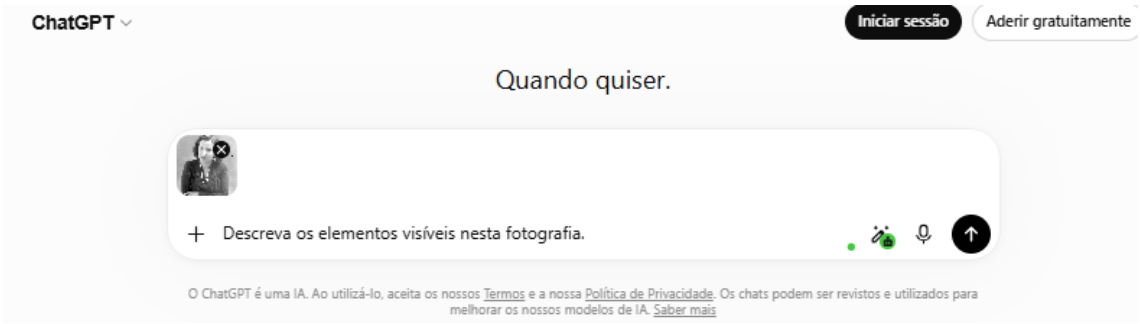



Imagem D.1: *Print* do ChatGPT Vision *input* Teste 1

Teste 1:*Output* (resposta gerada):



Descreva os elementos visíveis nesta fotografia.

Esta fotografia em preto e branco mostra uma pessoa posando para um retrato de estúdio. Posso descrever os elementos visíveis sem identificar quem ela é:

- A pessoa está sentada ou apoiada em uma cadeira, com os braços cruzados à frente do corpo.
- Tem cabelo escuro, curto a médio, penteado com uma risca lateral e cachos nas laterais.
- Usa brincos longos e um colar de pérolas com duas voltas.
- Veste uma blusa ou vestido escuro de mangas compridas, com detalhes bordados ou decorativos na parte frontal.
- Há anéis visíveis em pelo menos uma das mãos e uma pulseira no pulso direito.
- A expressão facial parece neutra a levemente atenta, com o olhar direcionado para um ponto fora da câmera.
- O fundo é simples e uniforme, típico de retratos de estúdio antigos.
- A imagem apresenta marcas de envelhecimento, como riscos, manchas, pontos de poeira e desgaste, sugerindo que se trata de uma fotografia antiga ou de uma digitalização de um negativo/fotografia histórica.

Pelo estilo do penteado, das joias, da roupa e pela técnica fotográfica, a imagem parece ser de meados do século XX, embora não seja possível determinar uma data exata apenas pela fotografia.

Imagem D.2: *Print output* do ChatGPT Vision, Teste 1 – Fotografia 1



Descreva os elementos visíveis nesta fotografia

Esta fotografia em preto e branco parece ser um retrato formal de estúdio mostrando duas pessoas lado a lado. Posso descrever os elementos visíveis sem identificar quem são:

- Há uma mulher à esquerda e um homem à direita, ambos em pé e voltados para a câmera.
- A mulher segura o braço esquerdo do homem, sugerindo uma pose de casal ou de familiares próximos.
- Ambos apresentam expressões neutras e postura ereta.

**Vestimenta da mulher**

- Usa um vestido escuro de mangas compridas e comprimento abaixo dos joelhos.
- O vestido possui uma gola clara, provavelmente branca ou de tom muito claro.
- Usa vários colares longos com medalhões ou pingentes grandes que descem até a cintura.
- Tem brincos visíveis.
- Calça meias escuras e sapatos fechados com detalhe decorativo na parte superior.
- O cabelo é curto e ondulado, penteado rente à cabeça.

**Vestimenta do homem**

- Usa um terno escuro de abotoamento duplo (duas fileiras de botões).
- Veste camisa clara e gravata.
- Há um lenço visível no bolso do paletó.
- Usa calças sociais largas e sapatos escuros.
- Tem cabelo curto, cuidadosamente penteado com risca lateral.
- Possui um pequeno bigode.

**Ambiente**

- O retrato foi feito em estúdio fotográfico.
- O fundo é pintado, representando uma paisagem com elementos arquitetônicos e vegetação.
- À esquerda é possível distinguir uma estrutura semelhante a uma torre ou edifício religioso.
- O piso do estúdio é visível na parte inferior da imagem.

**Estado da fotografia**

- A imagem apresenta riscos, manchas e sinais de desgaste típicos de fotografias antigas ou de negativos digitalizados.
- Pelo estilo das roupas, penteados e da fotografi estúdio, a cena parece datar aproximadamente das décadas de 1920 ou 1930, embora não seja possível determinar uma data exata apenas pela imagem.

Imagem D.3: *Print output* do ChatGPT Vision, Teste 1 - Fotografia 2



Descreva os elementos visíveis nesta fotografia

Esta fotografia em preto e branco mostra uma pessoa posando sozinha em um estúdio fotográfico. Eis os principais elementos visíveis:

**Pessoa retratada**

- A pessoa está em pé, voltada para a câmera.
- Tem o cabelo penteado para trás e preso junto à cabeça.
- A expressão é séria ou neutra.
- Usa brincos pequenos e um colar justo ao pescoço, além de um colar mais longo com dois medalhões ou pingentes circulares.

**Vestuário**

- Veste um traje escuro de mangas compridas.
- A gola é larga e decorada com bordados ou aplicações em padrão ornamental.
- Na região da cintura há uma espécie de avental ou sobreposição de tecido com detalhes rendados.
- Um pequeno lenço ou objeto de tecido aparece preso à frente da roupa, próximo ao quadril esquerdo.
- Usa meias escuras e sapatos fechados.

**Mobiliário e objetos**

- À direita da pessoa há um pedestal alto de madeira torneada, com uma coluna em espiral.
- Sobre o pedestal encontra-se um vaso decorativo.
- O vaso contém um arranjo de flores de hastes finas e pétalas claras, semelhantes a margaridas ou flores silvestres.
- A pessoa apoia levemente uma das mãos sobre o pedestal.

**Cenário**

- O fundo é um painel pintado típico de estúdios fotográficos antigos.
- A pintura sugere uma paisagem romântica ou de jardim, com:
  - uma escadaria com balaustrada;
  - vegetação e árvores;
  - uma construção ao fundo que lembra uma torre ou edifício religioso.

**Estado da fotografia**

- Há manchas, pontos, riscos e áreas desgastadas espalhadas pela imagem.
- O contraste relativamente suave e os sinais de deterioração indicam que se trata de uma fotografia antiga ou de uma reprodução de um negativo histórico.

Pelos estilos da roupa, dos acessórios, do cenário de estúdio e da técnica fotográfica, a imagem parece ter sido produzida aproximadamente entre as décadas de 1910 e 1930, embora não seja possível determinar uma data exata apenas pela observação visual.

Imagem D.4: *Print output* do ChatGPT Vision, Teste 1 - Fotografia 3

## Teste 2 -com orientação

*Teste 2: Input (prompt):*

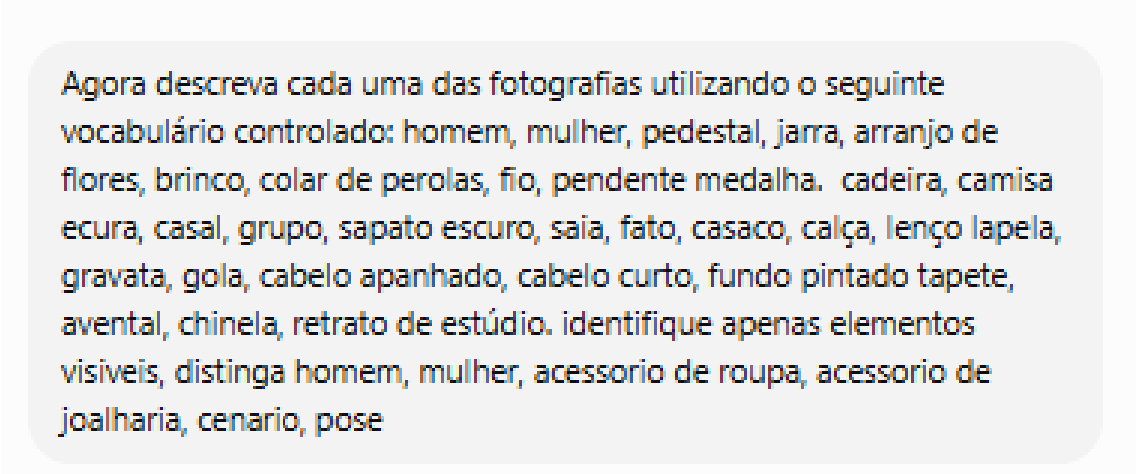



Imagem D.5: *Print input* do ChatGPT Vision, Teste 2

Teste 2: *Output* (resposta gerada):

A seguir descrevo cada fotografia utilizando apenas o vocabulário controlado fornecido, organizado por categorias.

### Fotografia 1

**Pessoa**

- mulher

**Pose**

- sentada
- apoiada em cadeira
- retrato de estúdio

**Acessórios de joalharia**

- brinco
- colar de perolas

**Acessórios de roupa**

- fio

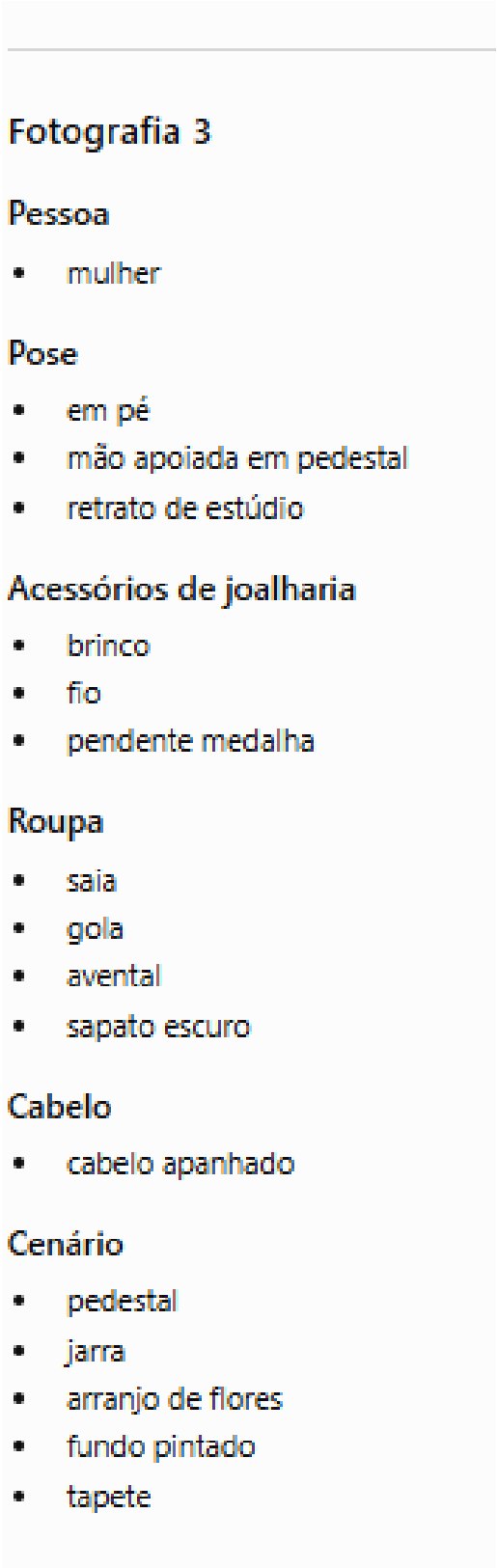
**Cabelo**

- cabelo curto

**Cenário**

- fundo pintado

Imagem D.6: *Print output* do ChatGPT Vision, Teste 2 - Fotografia 1



## Fotografia 3

### Pessoa

- mulher

### Pose

- em pé
- mão apoiada em pedestal
- retrato de estúdio

### Acessórios de joalharia

- brinco
- fio
- pendente medalha

### Roupa

- saia
- gola
- avental
- sapato escuro

### Cabelo

- cabelo apanhado

### Cenário

- pedestal
- jarra
- arranjo de flores
- fundo pintado
- tapete

Imagem D.7: *Print output* do ChatGPT Vision, Teste 2 - Fotografia 3

**Resumo por fotografia**

| Fotografia | Pessoas | Joalharia | Roupa | Cenário |
|---|---|---|---|---|
| 1 | mulher | brinco, colar de perolas | fio | cadeira, fundo pintado |
| 2 | casal (mulher + homem) | brinco, fio, pendente medalha | gola, saia, fato, casaco, calça, gravata, lenço lapela, sapato escuro | fundo pintado, tapete |
| 3 | mulher | brinco, fio, pendente medalha | gola, saia, avental, sapato escuro | pedestal, jarra, arranjo de flores, fundo pintado, tapete |

Imagem D.8: *Print output* do ChatGPT Vision, Teste 2 - Resumo