



ANA FILIPA SALADA SERÓDIO

***Gender Bias* em Inteligência Artificial Generativa**

Dissertação com vista à obtenção do grau de
Mestre em Direito na especialidade de Direito Internacional e Europeu

Orientadora:
Professora Doutora Helena Pereira de Melo, Professora da Faculdade de Direito da
Universidade Nova de Lisboa

Setembro 2024



ANA FILIPA SALADA SERÓDIO

Gender Bias em Inteligência Artificial Generativa

Dissertação com vista à obtenção do grau de
Mestre em Direito na especialidade de Direito Internacional e Europeu

Orientadora:
Professora Doutora Helena Pereira de Melo, Professora da Faculdade de Direito da
Universidade Nova de Lisboa

Setembro 2024

Declaração de Compromisso Antiplágio

Eu, Ana Filipa Salada Serôdio, declaro por minha honra que a Dissertação de Mestrado que apresento é original e que todas as minhas citações estão corretamente identificadas. Tenho consciência de que a utilização de elementos alheios não identificados constitui uma grave falta ética e disciplinar.

Lisboa, 15 de setembro de 2023

Ana Serôdio

Agradecimentos

Primeiramente, quero agradecer aos meus pais, que me apoiaram sempre e que me permitiram ter o mestrado que queria, e que mesmo com as nossas diferenças tentaram sempre compreender e ajudar da melhor forma que conseguiam. Nunca vou conseguir agradecer o suficiente por serem os melhores pais que podia ter. Quero agradecer ao meu irmão, que continua a ser uma inspiração para mim, e com quem criei a amizade mais especial que tenho. Agradeço à minha cunhada, por me incentivar sempre a sair da zona de conforto e a comprovar por mim própria que era capaz do que quisesse. Deixo também um agradecimento especial à minha afilhada Clarinha, que foi a maior alegria destes dois anos e a quem espero que este percurso académico inspire de alguma forma.

Agradeço aos meus avós, que não puderam estudar, mas que sempre nos deram valores e motivação para fazer diferente.

Quero agradecer às minhas amigas que perderam o seu tempo a rever esta tese e a dar-me força para continuar. Obrigada, Adriana, Ana, Leonor, Xana e Fátima!

Adriana, não me vou esquecer de todo o trabalho que tiveste, nem de todas as vezes que me incentivaste a não desistir, seja nesta ou em todas as fases que passámos juntas, onde sempre nos ajudámos a alcançar todos os objetivos, e em que festejámos as conquistas uma da outra como se fossem as nossas. Marta, obrigada por todo o apoio e amizade que tiveste para comigo durante estes anos, e por acreditares mais em mim do que eu própria.

Quero agradecer a todas as amigas da Barítuna, que de uma forma ou de outra me motivaram, animaram, e fizeram os meus dias mais felizes. Continuam sempre a ser o significado de casa e família.

Agradeço à ETUCE, especialmente à Martina, que me formou enquanto estagiária e que rapidamente se tornou uma inspiração a nível profissional.

Agradeço à Faculdade de Direito da Universidade Nova de Lisboa e à Faculdade de Direito da Universidade de Lisboa, por este percurso de grande crescimento académico e pessoal. Termino a agradecer à Professora Doutora Helena Pereira de Melo, por toda a disponibilidade prestada para me guiar neste percurso, e por todas as palavras de incentivo, seja quanto ao meu percurso académico ou ao meu futuro profissional. Foi muito importante ter alguém que admiro a acreditar em mim.

Modo de Citar

A presente Dissertação de Mestrado segue o novo acordo ortográfico da língua portuguesa, aprovado pela Resolução da Assembleia da República n.º 35/2008, de 29 de julho, com exceção das citações que mantemos a grafia utilizada por cada autor.

Em relação à bibliografia, vem ordenada alfabeticamente pelo último apelido do autor, seguindo a ordem: APELIDO, nome do autor, título da obra, editor, volume (se existir no caso concreto) e ano.

É de referir ainda que em relação aos documentos consultados informaticamente, é disponibilizado o sítio da Internet para consulta e indicado o dia em que foi consultado o mesmo.

Este trabalho possui uma lista de “Outras Fontes” que elenca a legislação citada, bem como artigos obtidos exclusivamente através do respetivo sítio da Internet.

As citações realizadas ao longo do relatório encontram-se entre aspas, com correspondente nota de rodapé apresentada da seguinte forma: NOME DO AUTOR, título da obra, editora, volume (se existir no caso concreto), ano e página. Nas citações em que se repete o mesmo livro algumas vezes, optámos por colocar a nota de rodapé da seguinte forma: NOME DO AUTOR, op. cit. e página.

As palavras em língua estrangeira encontram-se no tipo de letra itálico.

As citações em língua estrangeira encontram-se traduzidas no texto em português, e na língua original em nota de rodapé.

Por último, as abreviaturas e as siglas encontram-se identificadas também por ordem alfabética na “Lista de Abreviaturas e Siglas” na página seguinte.

Lista de Abreviaturas e Siglas

AAUW	American Association of University Women
<i>AI Act</i>	EU Artificial Intelligence Act
CADH	Convenção Americana sobre Direitos Humanos
CDD	Desigualdade Demográfica Condicional
CEDH	Convenção Europeia para a Proteção dos Direitos do Homem e das Liberdades Fundamentais
CES	Centro de Estudos Sociais
DUDH	Declaração Universal dos Direitos Humanos
FAMGA	Facebook, Apple, Microsoft, Google e Amazon
FATE	Fairness, Accountability, Transparency, and Ethics
FRIAs	Fundamental Rights Impact Assessments
GANs	Redes Neurais Generativas Adversariais
GEGI	Gender Employment Gap Index
IA	Inteligência Artificial
IAG	Inteligência Artificial Generativa
KOFAVIV	Comissão de Mulheres Vítimas para Vítimas
LLMs	Large Language Models
ODS	Objetivos de Desenvolvimento Sustentável
OIT	Organização Internacional do Trabalho
ONGs	Organizações Não Governamentais
PIDCP	Pacto Internacional sobre os Direitos Civis e Políticos
PIDESC	Pacto Internacional sobre os Direitos Económicos, Sociais e Culturais
RGPD	Regulamento Geral sobre a Proteção de Dados
RNNs	Redes Neurais Recorrentes
STEM	Science, Technology, Engineering and Mathematics
TFUE	Tratado sobre o Funcionamento da União Europeia
TIC	Tecnologias da Informação e Comunicação
TJUE	Tribunal de Justiça da União Europeia
TUE	Tratado da União Europeia
UE	União Europeia

VAEs	Autoencoders Variacionais
4IR	Fourth Industrial Revolution

Declaração de Conformidade

Declaro que esta Dissertação de Mestrado, incluindo espaços e notas de rodapé, ocupa um total de 159.483 caracteres.

Resumo

A tecnologia de Inteligência Artificial Generativa (IAG), que inclui modelos de linguagem e modelos de geração de imagens, tem sido amplamente utilizada em vários setores, desde a produção de conteúdo digital até o auxílio na tomada de decisões.

Contudo, com o avanço dessas tecnologias, surgem preocupações em relação à perpetuação de *bias* que podem ter impactos negativos na sociedade ao fortalecer estereótipos ou gerar desequilíbrios.

Na presente dissertação, é feita uma investigação sobre as origens e representações de *gender bias* em sistemas de IAG, examinando como esses modelos são treinados com grandes conjuntos de dados e como a representação desproporcional dos géneros nos dados de treino afeta os resultados. O estudo investiga as repercussões concretas de *gender bias* em situações comuns, como ferramentas de design, e modelos de linguagem. Além disso, são abordadas também as implicações éticas e sociais deste tema.

A fim de reduzir esses *biases*, são sugeridas soluções técnicas, como a seleção cuidadosa dos dados, a implementação de métodos justos de aprendizagem e a monitorização constante dos modelos. Para além disso, a presente dissertação analisa políticas e normas que podem contribuir para o desenvolvimento de uma IAG mais igualitária e justa.

Palavras-Chave: Igualdade de Género; Inteligência Artificial; Inteligência Artificial Generativa; Direito da União Europeia; Direito Internacional; Direitos Fundamentais; Carreira Profissional; Educação.

Abstract

Generative Artificial Intelligence (GenAI) technology, including large language models and image generation models, has been widely used in various sectors, from digital content production to decision-making.

However, with the advance of these technologies, concerns have arisen about the perpetuation of biases that can have negative impacts on society by strengthening stereotypes or generating imbalances.

This thesis investigates the origins and representations of gender bias in GenAI systems, examining how these models are trained with large data sets and how the disproportionate representation of genders in the training data affects the results. The study investigates the concrete repercussions of gender bias in common situations, such as design tools, and language models. In addition, the ethical and social implications of this issue are also addressed.

To reduce these biases, technical solutions are suggested, such as the careful selection of data, the implementation of fair learning methods and the constant monitoring of models. The present dissertation also analyses policies and norms that can contribute to the development of a more equal and fairer GenAI.

Keywords: Gender Equality; Artificial Intelligence; Generative Artificial Intelligence; European Union Law; International Law; Fundamental Rights; Professional Career; Education.

1. Introdução

A rápida expansão da inteligência artificial (IA) nos últimos anos tem transformado diversos setores, trazendo consigo inovações impressionantes e desafios significativos. Entre esses desafios, destaca-se a perpetuação e amplificação de estereótipos de gênero, que já eram problemáticos, em contextos sociais mais amplos, e que agora se manifestam de maneira preocupante nos algoritmos e sistemas de IA.

Esta dissertação explora e pretende analisar como é que esses estereótipos se infiltram na IA, particularmente na Inteligência Artificial Generativa (IAG), analisando as origens desse *bias* e as possíveis estratégias para mitigá-lo.

A disseminação de ferramentas de IAG, como o DALL-E 2 e ChatGPT, amplificou debates sobre os impactos sociais da IA. Grandes empresas como OpenAI, Google e Microsoft estão na vanguarda do desenvolvimento dessas tecnologias, e as discussões sobre a equidade e os efeitos sociais dessas ferramentas são mais relevantes do que nunca.

O entusiasmo com as capacidades inovadoras da IA vem acompanhado de uma preocupação crescente com a replicação e intensificação de preconceitos sociais existentes, incluindo estereótipos de gênero.

Líderes na área de IA, como GEOFFREY HINTON e SAM ALTMAN¹, expressaram preocupações sobre os riscos da IA, tanto em termos da sua evolução no futuro quanto aos danos atuais que já estão a ser causados, especialmente para grupos historicamente marginalizados e sub-representados. Sendo, por isso, necessário estudar e avaliar os danos relevantes nos dias de hoje.

Para combater essas desigualdades, é necessário adotar uma abordagem integrada que contemple tanto as mudanças tecnológicas como as sociais. A regulação da IA tem

¹ LAMENSCH, M., Generative AI Tools Are Perpetuating Harmful Gender Stereotypes, Centre for International Governance Innovation, disponível em: <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/>, consultado em 12.07.2024.

ganho cada vez mais atenção internacional, o que representa um ponto crucial nessa abordagem.

A União Europeia (UE) tem implementado medidas legais importantes para assegurar o respeito pela igualdade de género e não discriminação nas tecnologias em desenvolvimento. Por esse motivo, a presente dissertação irá analisar o papel regulatório da União Europeia no que diz respeito ao *AI Act*, cujo objetivo é assegurar a segurança, equidade e ausência de discriminação no desenvolvimento da IA, incluindo o *gender bias*.

O *gender bias* na IA é um reflexo das desigualdades sociais mais amplas que existem há vários anos. Esses padrões podem ser classificados em duas grandes categorias: estereótipos técnicos e estereótipos sociais.

Os estereótipos técnicos são o resultado de erros de algoritmos de aprendizagem supervisionada. Esses algoritmos podem produzir resultados tendenciosos que refletem as desigualdades presentes nos dados de treino. Por outro lado, os estereótipos sociais estão profundamente enraizados na estrutura social e são difíceis de corrigir porque refletem políticas, atitudes e preconceitos de longa data.

O *bias* na aprendizagem da máquina geralmente reflete as realidades tendenciosas e injustas da sociedade moderna. Para resolver estas questões, é essencial reconhecer que existem desigualdades sistêmicas perpetuadas por instituições, baseadas em desequilíbrios históricos. A luta contra esses estereótipos exige uma abordagem abrangente que inclua tanto mudanças tecnológicas quanto sociais.

A legislação da UE desempenha um papel crucial na luta contra a discriminação e na promoção da igualdade de oportunidades. Em matéria de não discriminação, a legislação da UE proíbe tanto a discriminação direta quanto a indireta.

A discriminação direta ocorre quando indivíduos são tratados de forma menos favorável com base em atributos protegidos, como a orientação sexual ou o género, enquanto a discriminação indireta envolve políticas ou práticas aparentemente neutras que resultam em desvantagens significativas para determinados grupos. Embora diferentes grupos recebam diferentes níveis de proteção, a obtenção da igualdade formal, baseada no princípio aristotélico de tratar casos semelhantes da mesma forma, não garante a igualdade de oportunidades.

A legislação da UE em matéria de não-discriminação, tal como interpretada pelo Tribunal de Justiça da União Europeia (TJUE), enfatiza a igualdade material como o seu objetivo pretendido. A lei apoia medidas de igualdade de oportunidades e ações positivas para eliminar sistematicamente as desigualdades ao longo do tempo.

O reconhecimento pelo TJUE da aplicabilidade horizontal do artigo 21.º da Carta dos Direitos Fundamentais sublinha o caráter geral e fundamental dos princípios de não discriminação na UE. Isto significa que os princípios de não discriminação se aplicam não apenas em contextos verticais, como nas relações entre os indivíduos e o Estado, mas também em contextos horizontais, por exemplo, as relações entre indivíduos e organizações privadas, incluindo empresas de tecnologia.

O conceito de “discriminação indireta” na legislação da UE é consistente com as noções modernas de justiça algorítmica. As justificações utilizadas pelo TJUE para medir medidas paralelas de discriminação indireta, como a igualdade demográfica, destacam a necessidade de uma representação justa entre grupos reconhecidos.

Ao percebermos que a IAG oferece oportunidades importantes para o progresso, também entendemos que esta representa riscos, tal como, perpetuar e exacerbar os estereótipos de género. Para mitigar estes riscos, é necessária uma abordagem multifacetada que inclua mudanças tecnológicas e sociais, em conformidade com as diretivas e leis existentes que promovam a igualdade e a antidiscriminação.

Para cumprirmos os objetivos a que nos propomos, abordando todos estes tópicos, a presente dissertação é dividida em vários capítulos que retratam o *gender bias* na IAG e as suas implicações regulatórias.

Começaremos com uma análise da regulamentação da igualdade de género, apresentando um panorama sobre como esta evoluiu ao longo do tempo e as principais iniciativas tomadas tanto no âmbito internacional quanto europeu. Seguidamente será dado destaque à regulação da IA, com foco em documentos essenciais como a Resolução para a IA das Nações Unidas, o *AI Act* da União Europeia e a Convenção-Quadro do Conselho da Europa relativa à IA.

Após essa análise jurídica, iremos examinar a relação entre a IAG e o *gender bias*. Serão abordadas as características e progresso da IAG e, além disso será feita uma análise crítica sobre como esse tipo específico de IA contribui para manter as desigualdades de género.

As desigualdades de gênero nos sistemas de IAG serão exploradas, com especial atenção para a falta constante de representação feminina nas áreas de Ciência, Tecnologia, Engenharia e Matemática (Science, Technology, Engineering and Mathematics - STEM). Será ainda feita uma análise das causas e consequências dessas desigualdades, além de serem discutidas as maneiras como a IA colabora para perpetuar tais disparidades.

Em seguida, será discutido o funcionamento interno dos algoritmos e serão examinados os processos pelos quais surge o *bias* algorítmico, além das limitações da transparência nos sistemas de IA.

Faremos também uma avaliação concreta dos impactos do *gender bias*, utilizando o exemplo de um estudo da UNESCO e um estudo de caso, que se encontra em anexo, realizado nesta pesquisa, utilizando sistemas de IAG de geração de imagem.

Após esta investigação, é analisado o impacto do *gender bias* na IAG e os efeitos que o mesmo tem na educação e, para além disso, na escolha da carreira profissional.

Finalmente, será elaborado um conjunto de sugestões de medidas para reduzir o *gender bias* nos sistemas de IAG, onde se incluem ideias que visam melhorar a transparência dos algoritmos, promover uma maior igualdade na IAG e fortalecer as regulamentações existentes.

Concluiremos a presente dissertação com uma análise sobre o impacto da IA na sociedade futura e a necessidade de assegurar que essas tecnologias emergentes sejam desenvolvidas seguindo princípios éticos e inclusivos, com ênfase em fomentar a igualdade de gênero.

2. Regulação da Igualdade de Género

2.1. Evolução Histórica

A preocupação em incluir a igualdade de género na UE começou formalmente com o Tratado de Roma, que estabeleceu a base legal para a igualdade salarial entre homens e mulheres. Esse tratado foi um marco, pois pela primeira vez um documento internacional reconheceu a necessidade de igualdade no meio laboral. Desde então, a UE tem implementado uma série de medidas legislativas e políticas para promover a igualdade de género.

Ao longo das décadas, a legislação europeia evoluiu para incluir a igualdade de tratamento das mulheres no acesso ao trabalho, formação, promoções, condições de trabalho, remuneração e prestações de segurança social. Além disso, direitos como a licença parental e a conciliação entre a vida pessoal e familiar foram garantidos, refletindo um compromisso contínuo com a igualdade de género.

Os tratados da UE, nomeadamente o Tratado da União Europeia (TUE) e o Tratado sobre o Funcionamento da União Europeia (TFUE), reforçam a igualdade de género como um objetivo essencial. O TUE consagra a igualdade entre mulheres e homens nos artigos 2.º e 3.º, n.º 3, enquanto o TFUE atribui à União a tarefa de eliminar as desigualdades e promover a igualdade de género em todas as suas atividades, conforme dispõe o seu artigo 8.º.

A Carta dos Direitos Fundamentais da UE, proclamada em 2000 e em vigor desde 2009, é outro pilar importante na promoção da igualdade. No seu “*Título III – Igualdade*”, a Carta, nomeadamente o artigo 21.º, proíbe a discriminação com base em *"sexo, raça, cor ou origem étnica ou social, características genéticas, língua, religião ou convicções, opiniões políticas ou outras, pertença a uma minoria nacional, riqueza, nascimento, deficiência, idade ou orientação sexual"*. Além disso, garante a igualdade entre homens e mulheres em todos os domínios, incluindo em matéria de emprego, trabalho e remuneração, conforme o disposto no artigo 23.º.

Para continuar a fazer progressos concretos em matéria de igualdade de género e alcançar os Objetivos de Desenvolvimento Sustentável (Agenda 2030), a Comissão Europeia, sob a liderança de URSULA VON DER LEYEN, lançou a Estratégia para a Igualdade de Género 2020-2025. Esta estratégia estabelece as principais ações a serem

realizadas nos próximos anos, centrando-se, nomeadamente em: erradicar a violência e os estereótipos; promover uma vida próspera e igualitária; equidade nos cargos de liderança; integração da perspectiva de género; financiamento para a igualdade de género; igualdade de género global.

2.2. Principais Iniciativas ao Nível Internacional

2.2.1. Carta das Nações Unidas

Antes de 1945, a discriminação era abordada principalmente em tratados sobre minorias, que tinham um âmbito limitado. Com a adoção da Carta das Nações Unidas, uma cláusula universal de não discriminação passou a fazer parte do direito internacional. Após as trágicas experiências da Segunda Guerra Mundial e, paralelamente do Holocausto, a ONU pretendia tornar-se um defensor global dos direitos individuais, aprendendo com as fraquezas da Liga das Nações. A Carta, quando redigida, sublinhava explicitamente a promoção dos direitos humanos, especialmente a igualdade e a não discriminação, como uma disposição fundamental.

As principais disposições da Carta das Nações Unidas relativas aos direitos humanos, como o nº 3 do artigo 1º, a alínea c) do artigo 55º e o artigo 56º, sublinharam o compromisso de proteção dos direitos humanos. Para além disso, outros artigos da Carta salientavam a missão da ONU de desenvolver relações amistosas baseadas na igualdade de direitos e na autodeterminação, promovendo o respeito pelos direitos humanos e pelas liberdades fundamentais sem distinção.

Apesar destas disposições, a Carta não estabeleceu imediatamente obrigações nem definiu “direitos humanos e liberdades fundamentais”. Em vez disso, delineou uma vaga obrigação de promover o respeito universal e a cooperação, ficando aquém, salvo melhor entendimento, do objetivo a que se propôs, sendo que a única disposição inequívoca era a proibição da discriminação, regulada no seu artigo 21.º.

2.2.2. Declaração Universal dos Direitos Humanos

No âmbito internacional dos direitos humanos, é garantido a todas as pessoas o direito à proteção igualitária contra qualquer forma de discriminação. Garantir que todos

possam usufruir dos direitos e liberdades estabelecidos, sem qualquer forma de distinção, é um princípio fundamental.

A Declaração Universal dos Direitos Humanos (DUDH) é um dos principais instrumentos que reforçam esse direito: todas as pessoas possuem a habilidade de desfrutar dos direitos e das liberdades estabelecidos nesta Declaração, sem qualquer tipo de discriminação, seja com base em raça, cor, sexo, língua, religião, opinião política ou qualquer outra origem nacional ou social. Isso também se aplica à propriedade. Além disso, é garantido a todos o direito de igual proteção perante a lei, sem qualquer tipo de discriminação. Também têm direito à mesma proteção contra discriminações que violem essa Declaração e qualquer incitação para tal ato, conforme o seu artigo 7º.

Um exemplo disso é quando os modelos generativos de IA produzem resultados que são depreciativos ou prejudiciais em relação a pessoas com identidades marginalizadas, promovendo estereótipos falsos e nocivos, além de contribuir para diversas formas de discriminação na sociedade. Os resultados dos modelos de IA generativa frequentemente favorecem grupos culturalmente hegemónicos, como homens brancos, ocidentais, masculinos, heterossexuais e cisgéneros. Isso pode levar à representação incorreta ou sub-representação de outros grupos em larga escala.

Além disso, os conjuntos de dados de treino da IA generativa costumam negligenciar as "línguas com recursos limitados", resultando numa baixa eficácia dos sistemas para os falantes desses idiomas. A IA generativa pode causar discriminação e aprofundar ainda mais a divisão digital existente entre países com recursos abundantes e países com recursos limitados, devido ao seu desempenho insatisfatório em contextos linguísticos, geográficos e culturais específicos.²

A DUDH desenvolve os princípios de igualdade da Carta das Nações Unidas, com os seus 30 artigos a abordarem explícita ou implicitamente a igualdade. Entre os artigos mais importantes destacamos o artigo 1.º, que declara que *“Todos os seres humanos*

² B-TECH, Taxonomy of Human Rights Risks Connected to Generative AI Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI., disponível em: <https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf>, consultado em 12.07.2024.

nascem livres e iguais em dignidade e em direitos. Dotados de razão e de consciência, devem agir uns para com os outros em espírito de fraternidade”, e o artigo 2.º. Os artigos 4º e 7º salientam ainda a proibição da escravatura e a igualdade de proteção perante a lei.

Não tão relevantes quanto a Carta e a DUDH, mas ainda assim consideramos importante referir, dois pactos internacionais: o Pacto Internacional sobre os Direitos Civis e Políticos (PIDCP, 1966) e o Pacto Internacional sobre os Direitos Económicos, Sociais e Culturais (PIDESC, 1966), sendo que o artigo 26.º do PIDCP proíbe explicitamente a discriminação, garantindo a igualdade e a eficácia da proteção contra a discriminação por diversos motivos. Para além disso, o PIDESC contém ainda cláusulas gerais e específicas sobre a não-discriminação.

Tratados específicos, como a Convenção Internacional sobre a Eliminação de Todas as Formas de Discriminação Racial (CERD, 1965) e a Convenção sobre a Eliminação de Todas as Formas de Discriminação contra as Mulheres (CEDAW, 1979), desenvolvem a não discriminação e sublinham o empenho da comunidade internacional na luta contra a discriminação.

As convenções regionais de direitos humanos, incluindo a Convenção Europeia para a Proteção dos Direitos do Homem e das Liberdades Fundamentais (CEDH, 1950) e a Convenção Americana sobre Direitos Humanos (CADH, 1969), incorporam cláusulas de não discriminação semelhantes, por exemplo, o artigo 14º da CEDH e o artigo 24º da CADH.

Para compreender o direito à igualdade e à não discriminação, é essencial clarificar o conceito de discriminação e a sua relação com a igualdade. A igualdade e a não discriminação são reconhecidas como expressões positivas e negativas do mesmo princípio, sendo que a igualdade implica a ausência de discriminação. A Subcomissão para a Prevenção da Discriminação e Proteção dos Direitos Humanos, criada pelas Nações Unidas, desempenhou um papel crucial na abordagem das questões da discriminação.

Os primeiros debates no seio da Subcomissão não conduziram a uma definição jurídica, mas sublinharam a prevenção de qualquer ação que negasse a indivíduos e/ ou a grupos a igualdade de tratamento. O tratamento diferenciado era considerado justificável quando servia a satisfação e o bem-estar da comunidade. Surgiu uma

distinção crucial entre a diferenciação justificada para uma verdadeira igualdade e a discriminação injustificada baseada em distinções indesejadas, irrazoáveis ou iníquas.

O PIDCP e o PIDESC, embora não forneçam definições explícitas de “discriminação”, centram-se em motivos específicos. O artigo 1º do CERD define a discriminação racial e o artigo 1º da CEDAW define a discriminação contra as mulheres. No entanto, o termo discriminação deve ser entendido de forma ampla, de modo a abranger distinções baseadas em vários motivos que afetam o reconhecimento, o gozo ou o exercício dos direitos e liberdades de todos os indivíduos.

O “conceito composto de discriminação” universal incorpora elementos como a estipulação de uma diferença de tratamento, a produção de um determinado efeito e o facto de se basear num motivo proibido. Os termos “distinção”, “exclusão”, “restrição” e “preferência” descrevem o tratamento diferenciado, e qualquer um destes termos pode estabelecer uma ação por discriminação. As preferências podem, por exemplo, envolver a promoção não razoável de um grupo em detrimento de outros.

O tratamento diferenciado, por si só, não é suficiente para estabelecer a discriminação. É essencial uma justificação baseada em objetivos legítimos e critérios razoáveis e objetivos. O Comentário Geral n.º 18 refere que a diferenciação é permitida se servir um objetivo legítimo e se os critérios de diferenciação forem razoáveis e objetivos.

Há quatro tratados de direitos humanos que contêm definições explícitas sobre a discriminação. Para além do CERD e da CEDAW, que foram mencionados acima, a Convenção n.º 111 da Organização Internacional do Trabalho (OIT), relativamente à discriminação em matéria de emprego e profissão (1958) declara que o termo “discriminação” inclui *“toda a distinção, exclusão ou preferência fundada na raça, cor, sexo, religião, opinião política, ascendência nacional ou origem social, que tenha por efeito destruir ou alterar a igualdade de oportunidades ou de tratamento em matéria de emprego ou profissão.”*

De acordo com a Convenção da UNESCO contra a Discriminação na Educação (1966), o termo “discriminação” inclui distinção, exclusão, limitação ou preferência que, sendo baseada na raça, cor, sexo, língua, religião, opinião política ou outra, origem nacional ou social, condição económica ou nascimento, tenha o objetivo ou o efeito de anular ou prejudicar a igualdade de tratamento na educação.

Todas as quatro convenções, supramencionadas, referem-se ao “efeito” desse tratamento diferenciado. Três delas, com exceção da Convenção da OIT, consideram que a discriminação se baseia no “objetivo ou efeito”. O “objetivo” pode ser inferido como contendo um significado de “intenção”.

Por outro lado, a Convenção n.º 111 da OIT refere-se apenas a “efeito”, omitindo o conceito de “objetivo”, enquanto as outras três convenções utilizam as palavras “objetivo ou efeito”. A utilização da palavra “ou” em vez de “e” indica que o “objetivo” pode ser desvalorizado em comparação com o “efeito”. Uma vez que o conceito de finalidade contém um significado de intenção, é difícil definir e provar a intenção subjetiva necessária para estabelecer um ato discriminatório.

Consequentemente, a intenção discriminatória não é um elemento necessário da discriminação. A ênfase no “efeito” da ação e não na intenção significa que as medidas neutras serão consideradas “discriminatórias” se afetarem negativamente um grupo da sociedade que tenha sido escolhido para proteção. Uma tendência central no desenvolvimento da legislação sobre discriminação em muitas jurisdições nacionais tem sido a passagem de um requisito de intenção, para fundamentar uma queixa, para o reconhecimento do efeito adverso da discriminação.

Inicialmente, a responsabilidade por discriminação era circunscrita de forma muito restrita, exigindo uma forma de intenção que era equivalente a malícia ou má-fé. Atualmente, a legislação em matéria de discriminação tende a oscilar de um extremo para o outro, passando de um enfoque exclusivo na culpabilidade moral do arguido, para um enfoque exclusivo nos efeitos da discriminação nas suas vítimas. A palavra “discriminação”, por si só, é atualmente utilizada no sentido pejorativo, como sendo uma distinção injusta, não razoável, injustificável ou arbitrária.

O significado mais óbvio de discriminação realça a hostilidade ou o preconceito, mas deve ser adotada uma definição mais ampla: em primeiro lugar, porque os dados sugerem que o tratamento diferencial desvantajoso ocorre frequentemente na ausência de preconceito ou hostilidade e, em segundo lugar, devido à dificuldade inerente à definição ou prova de preconceito ou hostilidade.

No Reino Unido, por exemplo, uma das reações mais comuns às alegações nacionais de discriminação é: “Como se pode provar isto?”, e o pressuposto é que a intenção discriminatória deve ser um elemento essencial do erro. A prova da discriminação tem

três elementos: em primeiro lugar, temos de saber como é definida a discriminação no contexto jurídico em que supostamente surge. Em segundo lugar, temos de identificar o que deve ser provado para estabelecer a ocorrência de discriminação. Aqui, coloca-se a questão da intenção. E, por último, há a questão da obtenção das provas necessárias.

2.3 Regulação da Igualdade de Género na União Europeia

A legislação europeia sobre a não discriminação está presente tanto no direito primário quanto no secundário. O artigo 21.º da Carta dos Direitos Fundamentais da UE, que faz parte do direito primário, estabelece a proibição de qualquer discriminação baseada em fatores como sexo, raça, cor, origem étnica ou social, características genéticas, língua, religião ou crença, opinião política ou outra, pertença a uma minoria nacional, propriedade, nascimento, deficiência, idade ou orientação sexual. Esta lista não é exclusiva e é neutra em relação ao setor. No entanto, a Carta aplica-se apenas a órgãos públicos da União Europeia e dos Estados-Membros, não ao setor privado. Em contraste, a legislação secundária aplica-se tanto ao setor privado quanto ao público.

Como acontece em todas as Diretivas da UE, por oposição aos Regulamentos da UE, as Diretivas sobre a não discriminação estabelecem apenas um padrão mínimo e fornecem um quadro geral, sendo que este último tem de ser transposto para a legislação nacional pelos Estados-Membros.

Muitos Estados-Membros oferecem uma proteção maior do que os padrões mínimos estabelecidos pelas Diretivas. Ora, esta variabilidade resulta num padrão fragmentado de proteção contra a discriminação em toda a Europa.

A UE promove a integração da perspectiva de género através da sua estratégia de emprego, incluindo a igualdade de género como o quarto pilar das orientações para o emprego, juntamente com a empregabilidade, a adaptabilidade e o espírito empresarial. Para além disso, a UE exige que os Estados-Membros considerem o impacto do género em todas as políticas, no âmbito de cada um destes pilares, dando um grande impulso à integração das questões da igualdade de oportunidades no quadro do emprego. Esta abordagem faz parte da adoção pela UE do “método aberto de coordenação”, que permite aos Estados-Membros desenvolverem programas políticos adequados à sua situação específica, mas de acordo com orientações comuns acordadas.

Um dos objetivos para o desenvolvimento sustentável, a serem atingidos até 2030 pelas Nações Unidas, é a igualdade de género, sendo este o quinto objetivo. Para este ser cumprido são importantes os seguintes passos: o uso da tecnologia para promover a capacitação das mulheres, a adoção e fortalecimento de políticas e a legislação para promover a igualdade de género para todas as mulheres.

A UE tem demonstrado o seu empenho na igualdade de género nas suas políticas através de vários exemplos específicos, nomeadamente na promoção da integração da perspetiva de género na política de emprego e na política governamental em geral, tendo quase todos os Estados-Membros criado mecanismos formais de integração da perspetiva de género até 2001, conforme a Estratégia Europeia de Emprego. Além disso, foram estabelecidas metas ou quotas mínimas para a participação das mulheres nas políticas ativas do mercado de trabalho em vários países, como Alemanha, França, Bélgica e Áustria.

Os fatores que contribuem para moldar a legislação da UE em matéria de igualdade entre homens e mulheres incluem as aspirações da União Europeia em matéria de direitos humanos, compromissos de desenvolvimento sustentável e considerações demográficas. Outro fator é o alargamento do âmbito da igualdade entre homens e mulheres ao acesso e fornecimento de bens e serviços, bem como ao Terceiro Pilar, de acordo com o Tribunal de Justiça Europeu.

Esforços têm sido feitos para tornar visível e acessível o princípio da igualdade entre homens e mulheres, apesar de algumas disposições ainda serem pouco claras e contraditórias, salvo melhor entendimento.

O envolvimento de agentes externos na salvaguarda e no reforço do acervo comunitário em matéria de igualdade entre homens e mulheres na UE é evidente através de vários meios. As organizações da sociedade civil, os parceiros sociais e outros agentes foram incentivados a participar nas atividades da UE relacionadas com a igualdade entre homens e mulheres.

Para além disso, o Acordo de Política Social, anexo ao Tratado da União Europeia, facilitou o reforço do papel dos parceiros sociais no processo legislativo da Comunidade. Os parceiros sociais podem ser consultados sobre as ações comunitárias e o conteúdo das propostas legislativas, negociar acordos coletivos e participar no processo de elaboração de políticas. Adicionalmente, o envolvimento da sociedade civil

na UE, em especial no contexto da Convenção, tem sido elogiado e incentivado como uma forma de democratização da União.

A criação de novas agências para a igualdade e o envolvimento de associações, organizações e/ ou outras entidades jurídicas, na instauração de processos ou na atuação, em nome das vítimas de discriminação, exemplificam ainda mais o papel dos agentes externos na salvaguarda e no reforço da igualdade entre homens e mulheres na UE. Embora o envolvimento destes agentes externos seja em grande parte voluntário, tem contribuído para o reforço da proteção das vítimas de discriminação e para a promoção da igualdade entre homens e mulheres.

Não obstante, existem disposições pouco claras e contraditórias na legislação da UE em matéria de igualdade entre homens e mulheres, um exemplo claro disso é a cláusula geral sobre a antidiscriminação da Carta, que não distingue entre os vários motivos de discriminação, sendo que apenas proíbe a discriminação em sentido amplo.

A Diretiva n.º 2006/54/CE constitui um marco na dedicação da UE à erradicação da discriminação baseada no género no local de trabalho. A Diretiva garante salário igual para trabalho igual e proíbe a discriminação baseada no género no emprego, promovendo, ao mesmo tempo, o equilíbrio entre a vida profissional e a vida privada e protegendo os indivíduos contra o assédio e a violência baseados no género.

Simultaneamente, a Diretiva n.º 2004/113/CE do Conselho complementa os esforços para combater a discriminação entre homens e mulheres, abordando o acesso a bens e serviços e o seu fornecimento. Esta Diretiva proíbe a discriminação baseada no género no acesso a bens e serviços e no seu fornecimento, em especial, no setor dos seguros e dos serviços financeiros, abordando práticas não discriminatórias em áreas como a fixação de preços e a cobertura. A Diretiva permite exceções específicas e ações positivas que promovam a igualdade entre homens e mulheres, tendo sempre em conta requisitos como a proporcionalidade e a temporalidade.

O artigo 21.º da Carta dos Direitos Fundamentais reforça o empenho da UE na não discriminação, proibindo explicitamente a discriminação com base em vários motivos,

para além do género, criando uma salvaguarda abrangente contra todas as formas de discriminação na UE.³

No contexto da União, o princípio da não discriminação é crucial quando se trata do desenvolvimento e aplicação de tecnologias baseadas em Inteligência Artificial. Frente aos desafios apresentados pelos *biases* nos sistemas de inteligência artificial, surgem ideias para harmonizar essas tecnologias com as leis vigentes, procurando assegurar que a discriminação seja reduzida ao mínimo e que seja adequadamente analisada.

Entre as sugestões, encontra-se o uso do padrão estatístico conhecido como Desigualdade Demográfica Condicional (CDD)⁴. Foi criado este padrão com o objetivo de estar em conformidade com os critérios estabelecidos pelo TJUE para a avaliação de casos de discriminação segundo a legislação da UE.

O CDD desempenha o papel de um conjunto mínimo de evidências estatísticas em situações que envolvam discriminação automatizada, criando assim uma estrutura sólida para a sua rápida avaliação.

Ao seguir os padrões legais estabelecidos, torna-se mais fácil fazer avaliações precisas sobre eventuais preconceitos nos sistemas de IA, auxiliando na compreensão dos casos de discriminação. A proposta também salienta a relevância da regularidade dos procedimentos, sugerindo a implementação de protocolos padronizados para assegurar que os princípios de justiça sejam aplicados de maneira consistente. Essa medida possibilitaria que os tribunais tivessem flexibilidade para interpretar os casos, com base no contexto específico, de forma a incorporar considerações de justiça nos sistemas de IA e permitir diferentes interpretações judiciais.

Ao fornecer um quadro estatístico claro, o CDD não apenas auxilia juízes e profissionais do direito, mas também incentiva a colaboração entre programadores de tecnologia e especialistas jurídicos para impulsionar, simultaneamente, os avanços no campo tecnológico e no campo legal. Entretanto, o reconhecimento de que o *bias* não pode ser completamente eliminado dos sistemas de IA é essencial. O quadro proposto admite níveis aceitáveis de *bias*, estabelecendo limites claros para o que seria

³MASSELOT, A., The State of Gender Equality Law in the European Union. *European Law Journal*, 13(2), 2007, págs. 152–168.

⁴ WACHTER, S., MITTELSTADT, B., & RUSSELL, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, *SSRN Electronic Journal*, 2020, pág.6.

considerado ilegal, enquanto promove a responsabilização entre programadores e utilizadores de IA.

É de ressaltar a complexidade de alcançar sistemas de IA totalmente justos, enfatizando que o *bias* é uma questão dinâmica e contextual, influenciada por dados do mundo real, sendo certo, que isso exige avaliações contínuas para garantir a conformidade com os padrões legais à medida que os sistemas evoluem.

A flexibilidade judicial é igualmente importante, pois a determinação de limites aceitáveis para o *bias* é uma questão que deve ser resolvida pelos tribunais, sendo que estes têm de adaptar as leis existentes às nuances dos sistemas de IA.

As normas sobre a não discriminação da UE são amplas o suficiente para permitir a adaptação a novas tecnologias, e o poder judiciário tem a capacidade de interpretar essas leis de maneira a que se mantenha a justiça num cenário tecnológico em rápida mudança. Existe uma necessidade de colaboração interdisciplinar para desenvolver métricas de justiça completas que levem em conta as complexidades técnicas e sociais envolvidas.

O quadro proposto procura equilibrar os interesses das pessoas que desenvolvem a IA, utilizadores e indivíduos afetados, promovendo práticas éticas na implementação de IA e assegurando que os sistemas sejam justos e responsáveis. Ao adotar estas estratégias, a proposta visa estabelecer uma abordagem abrangente para mitigar o *bias* nos sistemas de IA, garantindo, ao mesmo tempo, a conformidade com o direito da União Europeia, protegendo os princípios da não discriminação, pois estes são fundamentais.

A distinção fundamental entre abordagens jurídicas tradicionais e os novos desafios impostos pela tecnologia é refletida pelas principais diferenças entre o direito da não discriminação da União Europeia e os conceitos de justiça algorítmica. O primeiro tem por base contextos históricos e sociais, com o objetivo principal de proteger indivíduos contra qualquer forma de discriminação fundamentada em características específicas, tais como a etnia e o género. Por outro lado, a justiça algorítmica tem como objetivo resolver o problema da discriminação que surge de sistemas automatizados e do processamento de dados, dando prioridade à análise de *biases* e equidade com base em medidas estatísticas.

Embora o direito à não discriminação da UE dependa significativamente da análise contextual e exija que os tribunais interpretem cada caso com base nas especificidades

das partes envolvidas, a justiça algorítmica prioriza a aplicação de modelos estatísticos e métricas para avaliar a equidade dos resultados. A análise jurídica proporciona uma profundidade contextual que, muitas vezes, falta nessa abordagem técnica.

Adicionalmente, o direito da UE permite que os juízes usem a discricionariedade nas suas decisões, o que possibilita diferentes interpretações normativas e reflexos das particularidades de cada caso. Em contrapartida, a justiça baseada em algoritmos costuma favorecer abordagens padronizadas para garantir uma avaliação consistente dos algoritmos.⁵

⁵ WACHTER, S., MITTELSTADT, B., & RUSSELL, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, SSRN Electronic Journal, 2020, págs.54-55.

3. Regulação da Inteligência Artificial

3.1. Visão Geral

A regulação da Inteligência Artificial é um tema que tem vindo a ganhar cada vez mais relevância e destaque na atualidade. À medida que a IA se torna progressivamente integrada em diversos setores da sociedade, as preocupações sobre a ética da sua utilização também aumentam. Em particular, a capacidade da IA de perpetuar e até exacerbar preconceitos existentes, como a discriminação de género, sendo este um dos principais desafios enfrentados pelos legisladores.

Este capítulo visa explorar as abordagens regulatórias adotadas por diferentes entidades, como a União Europeia e as Nações Unidas, e analisar a sua eficácia na mitigação do preconceito de género em sistemas de IA.

A análise da regulação da IA é essencial para a nossa dissertação, pois as políticas e regulamentações estabelecem o quadro dentro do qual estas tecnologias operam, sendo que as regulamentações determinam as obrigações dos programadores e operadores de IA para garantir que os seus sistemas não discriminem com base em género.

Ao entender as falhas e sucessos das atuais políticas regulatórias, podemos identificar quais as áreas onde estas são necessárias para melhor proteger os direitos das mulheres e de outros grupos marginalizados.

3.2. Resolução para a Inteligência Artificial das Nações Unidas

A 21 de março de 2024 a Assembleia-Geral das Nações Unidas adotou por consenso a sua primeira Resolução sobre a Inteligência Artificial⁶. Esta Resolução destaca os potenciais benefícios dos sistemas de IA que sejam seguros e fiáveis para o desenvolvimento sustentável, reafirmando os princípios do direito internacional e as resoluções anteriores relativas ao desenvolvimento sustentável e aos avanços tecnológicos. Patrocinada por vários países, incluindo Portugal, a Resolução sublinha ainda a importância de tirar partido da IA para alcançar os Objetivos de Desenvolvimento Sustentável (ODS).

⁶ NAÇÕES UNIDAS, ONU adota por consenso resolução para reger a inteligência artificial, ONU News, 2024, disponível em: <https://news.un.org/pt/story/2024/03/1829446>, consultado a 02.06.2024.

A Resolução evidencia a importância de eliminar o fosso digital entre géneros, que se refere às disparidades no acesso às tecnologias entre géneros. São incentivadas medidas específicas para melhorar o acesso, a literacia digital, a privacidade e a segurança em linha, em especial para as mulheres e para os grupos marginalizados. Esta abordagem visa garantir que as mulheres possam utilizar eficazmente as tecnologias, incluindo os sistemas de Inteligência Artificial, promovendo assim a igualdade de género no domínio digital.

A Resolução apela também à integração das perspetivas de grupos minoritários e marginalizados, nas decisões e nos quadros políticos. Ao incluir estas considerações, pretende-se criar políticas que não só reconheçam, mas também abordem ativamente, as barreiras enfrentadas pelas mulheres e outros grupos marginalizados no acesso às tecnologias digitais e na obtenção de benefícios.

O incentivo à investigação e à cooperação internacional é um aspeto significativo da Resolução. Isto inclui iniciativas que identifiquem e avaliem os impactos dos sistemas de Inteligência Artificial na igualdade de género. Ao promover a investigação inclusiva, a Resolução visa garantir que o desenvolvimento e a implementação de tecnologias digitais tenham em conta os impactos e benefícios específicos de género, promovendo assim avanços equitativos.

Para além disso, a Resolução salienta a necessidade de uma maior participação das mulheres nos setores da tecnologia e da IA. Ao melhorar as infraestruturas digitais, especialmente nos países em desenvolvimento, a Resolução apoia a capacitação das mulheres para participarem mais ativa e eficazmente na inovação tecnológica e na *governance*.

Esta primeira Resolução reconhece que a conceção e a utilização inadequadas de sistemas de IA podem reforçar as desigualdades e os preconceitos estruturais existentes. Salienta também a necessidade de criar meios de proteção para evitar a discriminação, refletindo assim o compromisso de promover a igualdade de género como parte de iniciativas mais amplas de justiça social.

Por último, para atingir estes objetivos, a Resolução defende o aumento do financiamento e do apoio a iniciativas de reforço de capacidades destinadas aos países em desenvolvimento, o que inclui a prestação de assistência técnica para garantir um acesso justo às tecnologias digitais, fator crucial para promover a igualdade de género

nestas regiões. Este apoio ajudará a capacitar as mulheres, reforçando as suas competências e oportunidades tecnológicas e digitais.⁷

3.3. *AI Act*⁸ - Regulamento para a Inteligência Artificial da União Europeia

O "EU Artificial Intelligence Act" (*AI Act*) é uma legislação fundamental, porquanto procura garantir a igualdade de género em sistemas de Inteligência Artificial.

O processo de aprovação do *AI Act* foi longo e envolveu muitas negociações, especialmente após o lançamento do ChatGPT no final de 2022. O Parlamento Europeu aprovou, formalmente, o *AI Act* a 13 de março de 2023, com uma larga maioria de votos.

A 21 de Maio de 2024, o Conselho da União Europeia aprovou formalmente o Regulamento, o qual foi publicado a 12 de julho de 2024 no Jornal Oficial da União Europeia, e entrou em vigor a 1 de agosto de 2024. Os requisitos deste Documento só serão totalmente aplicados daqui a dois anos, iniciando-se esse processo gradual de aplicação a 2 de fevereiro de 2025. A partir de dia 2 de agosto de 2026 o *AI Act* deverá ser aplicado na integralidade (artigo 13.º do *AI Act*).

A legislação visa harmonizar as regras para a Inteligência Artificial, protegendo os direitos fundamentais dos cidadãos da União Europeia e garantindo um desenvolvimento seguro e fiável de sistemas de Inteligência Artificial no mercado europeu, tanto no setor público como no privado.

O Regulamento classifica sistemas de Inteligência Artificial de alto risco e proíbe certas aplicações, enquanto os sistemas de risco limitado estão sujeitos a obrigações de transparência. Os sistemas de alto risco devem ser autorizados e cumprem mais obrigações regulamentares, incluindo a proibição de manipulação cognitiva. Para além disso, estes sistemas de IA classificados como "de alto risco" devem cumprir obrigações específicas para garantir que não têm um impacto negativo nos direitos fundamentais, incluindo a igualdade de género.

⁷ NAÇÕES UNIDAS, Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, 2024, ONU, disponível em: <https://digitallibrary.un.org/record/4040897?v=pdf>, consultado em 23.05.2024.

⁸ EUROPEAN COMMISSION, The Artificial Intelligence Act, 2021, disponível em: <https://artificialintelligenceact.eu/the-act/>, consultado em 26.07.2024.

Não obstante ter suscitado tanto elogios, bem como críticas, a União Europeia é pioneira na adoção de regulamentação abrangente para Inteligência Artificial, sendo certo que a versão final aprovada recebeu maior apoio.

Para monitorar a aplicação do *AI Act* serão criados: um *AI Office* na Comissão Europeia, um painel científico de peritos independentes, um *AI Board* com representantes dos Estados membros e um fórum consultivo de *stakeholders*. A proposta inicial foi apresentada em 2021, mas passou por várias alterações no Parlamento Europeu.

Um dos principais objetivos do *AI Act* é proteger os direitos fundamentais das pessoas, garantindo transparência e imparcialidade. Para evitar a violação de qualquer direito fundamental, o *AI Act* propõe a realização de "*bias audits*", que são uma espécie de auditorias feitas para identificar e corrigir discriminações de gênero que possam afetar direitos fundamentais presentes nos sistemas de IA, conforme o disposto no seu artigo 10.º, n.º 2.

Além disso, o *AI Act* exige que sejam feitas avaliações de impacto sobre os direitos fundamentais⁹, em que, nos termos do seu artigo 27.º, os fornecedores de sistemas de IA de alto risco são obrigados a realizar avaliações de impacto sobre os direitos fundamentais para analisar potenciais impactos nesses direitos, que incluem a avaliação dos riscos relacionados com a discriminação de gênero.

O requisito de conduzir avaliações de impacto sobre os direitos fundamentais aplica-se especificamente a órgãos públicos e a todos aqueles que prestam serviços públicos, mas não a todos os sistemas de IA, i.e., empresas privadas que implantam IA para recrutamento podem não estar sujeitas a essa obrigação.

O artigo 85.º concede às pessoas lesadas o direito de apresentar queixa se acreditarem que os seus direitos, ao abrigo do Regulamento, foram violados, o que pode incluir a discriminação de gênero. Além disso, a Comissão Europeia terá a capacidade de atualizar as categorias de sistemas de IA de alto risco, incluindo aqueles que podem afetar de forma desproporcional a igualdade de gênero.

⁹ Texto original: Fundamental Rights Impact Assessments (FRIAs).

A criação de autoridades nacionais competentes e do AI Office destina-se a supervisionar o cumprimento e a aplicação destas disposições para garantir a proteção contra a discriminação algorítmica com base no género. O AI Office terá ainda um papel importante no que toca ao crescimento da consciencialização sobre os novos riscos que as tecnologias de IA podem trazer, especialmente no que se refere à discriminação de género.

O *AI Act* está alinhado com esforços globais para regulamentar a IA, procurando abordar a igualdade de género e outras formas de discriminação em colaboração com organizações internacionais, como as Nações Unidas. Este mecanismo incentiva ainda a participação de diferentes partes interessadas, incluindo a sociedade civil, para garantir que diversas vozes e perspetivas sejam ouvidas durante o desenvolvimento e a implementação de sistemas de IA na sociedade civil.

Por outro lado, a aplicação destas disposições depende, em grande medida, dos Estados-Membros, o que exige a cooperação entre os organismos de controlo da não discriminação e da IA relevantes para salvaguardar a igualdade de género. A monitorização contínua dos sistemas de IA será obrigatória para garantir que as normas de igualdade de género e não discriminação sejam sempre respeitadas.

O *AI Act* adota também uma abordagem baseada no risco, com regras mais rigorosas para sistemas potencialmente prejudiciais. A nova legislação prevê multas e penalizações para incumprimentos, definidas como uma percentagem da faturação anual das empresas. Foram introduzidas medidas para assegurar a inovação, como ambientes de teste com acesso a condições do mundo real, mas com proteções específicas, também conhecidos como *sandboxes*.

Existem várias salvaguardas e obrigações previstas no *AI Act*, por exemplo, para a Inteligência Artificial de uso geral, há requisitos de transparência, incluindo o respeito pela legislação da União Europeia sobre direitos de autor e a publicação de informações detalhadas sobre os conteúdos usados no treino. Por outro lado, há limites para a utilização de sistemas de identificação biométrica pelas autoridades policiais, onde a identificação biométrica à distância "em tempo real" e "em diferido" está sujeita a autorizações judiciais e rigorosas condições de utilização. A classificação social e a Inteligência Artificial utilizada para manipular ou explorar vulnerabilidades dos utilizadores estão proibidas. Os consumidores têm o direito de apresentar queixas e a

receber explicações significativas sobre decisões de Inteligência Artificial que afetem os seus direitos

O *AI Act* proíbe aplicações de Inteligência Artificial que ameacem os direitos dos cidadãos, como, por exemplo, sistemas de categorização biométrica baseados em características sensíveis, reconhecimento de emoções no local de trabalho e nas escolas, e policiamento baseado exclusivamente em perfis ou características pessoais.

DRAGOȘ TUDORACHE¹⁰ afirmou que a União Europeia ligou o conceito de Inteligência Artificial aos valores fundamentais que formam a base das nossas sociedades, mas que ainda há muito trabalho pela frente que vai além do próprio *AI Act*.¹¹ A IA levar-nos-á a repensar o contrato social no centro das nossas democracias.

O Regulamento para a Inteligência Artificial vem responder, diretamente, às propostas dos cidadãos durante a Conferência sobre o Futuro da Europa, que durou um ano, de abril de 2021 a maio de 2022 incluindo o reforço da competitividade da União Europeia em setores estratégicos, uma sociedade segura e fiável, a promoção da inovação digital com supervisão humana e uma utilização responsável da Inteligência Artificial, e a melhoria do acesso dos cidadãos à informação.

3.3.1. Análise crítica do *AI Act*

É alvo de crítica a abordagem do *AI Act* em relação à classificação de sistemas de IA de alto risco no contexto da discriminação de género, destacando várias preocupações importantes.

A classificação de sistemas de IA como de alto risco é criticável, pois aciona a conformidade obrigatória com as salvaguardas do *AI Act*, destinadas a proteger os direitos fundamentais. No entanto, é de notar que certos sistemas de IA podem escapar a essa classificação devido a definições ambíguas, correndo o risco de comprometer as proteções contra a discriminação de género. Essa ambiguidade pode permitir que

¹⁰ Relator da Comissão das Liberdades Cívicas.

¹¹ PARLAMENTO EUROPEU, Regulamento Inteligência Artificial: Parlamento aprova legislação histórica, disponível em: <https://www.europarl.europa.eu/news/pt/press-room/20240308IPR19015/regulamento-inteligencia-artificial-parlamento-aprova-legislacao-historica>, consultado em 05.05.2024.

sistemas de IA que deveriam ser classificados como de alto risco, devido ao seu potencial de discriminação de género, sejam classificados de outra forma, não sendo, assim, sujeitos aos requisitos rigorosos que poderiam mitigar tal discriminação.

Além disso, observa-se que o *AI Act* não impõe avaliações obrigatórias de impacto sobre os direitos fundamentais e “*bias audits*” em todos os sistemas de IA. Essa limitação é particularmente preocupante para empresas privadas que utilizam sistemas de IA para recrutamento, as quais não estão cobertas por tais obrigações. Esta lacuna pode levar à perpetuação de preconceitos de género em sistemas de IA utilizados tanto no âmbito laboral, como noutras áreas críticas, uma vez que esses sistemas podem não passar por uma análise minuciosa para identificar e mitigar potenciais discriminações de género.

Discute-se também como é que certos critérios, referidos no *AI Act*, podem levar à exclusão de sistemas de IA da categoria de alto risco, mesmo que possam potencialmente contribuir para a discriminação de género. Pensamos, salvo melhor entendimento, que uma definição mais precisa do que constitui um sistema de IA de alto risco e como as disposições do *AI Act* se aplicam a diferentes tipos de sistemas de IA, particularmente aqueles que podem impactar a igualdade de género, poderiam levar a uma maior proteção contra a discriminação de género.

Lütz¹² sugere a incorporação de requisitos obrigatórios para sistemas de IA de baixo risco, semelhantes aos dos sistemas de alto risco, a adoção de práticas de design diversas e a adesão a diretrizes éticas para promover a igualdade de género no desenvolvimento de IA.

Além disso, o mesmo autor enfatiza a necessidade de monitoramento contínuo e obrigações de transparência abrangentes para garantir que os princípios de igualdade de género sejam mantidos. Na sua perspetiva, partilhada por nós na presente dissertação, podemos destacar a abordagem do *AI Act* em relação à classificação de sistemas de IA de alto risco, referindo as suas potenciais deficiências em abordar adequadamente a discriminação de género, bem como a falta de definições mais claras, um alcance mais

¹² LÜTZ, F., The AI Act, gender equality and non-discrimination: what role for the AI office?, ERA Forum 25, 2024, págs. 79–95.

amplo e requisitos mais rigorosos. Com estas imprecisões, o *AI Act* pode falhar em prevenir preconceitos de género em sistemas de IA, ou, pelo menos, em preveni-los de forma integral.

Em conclusão, embora o *AI Act* introduza vários mecanismos destinados a abordar os preconceitos de género nos sistemas de IA, a sua eficácia pode ser limitada pelo âmbito da sua aplicação e pela necessidade de várias melhorias regulamentares. A aposta no equilíbrio entre os géneros nas equipas de desenvolvimento, as avaliações obrigatórias e as auditorias de discriminação são passos positivos para a promoção da igualdade de género nas tecnologias de IA.

3.4. Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial

A 5 de setembro de 2024, a Vice-Presidente da Comissão Europeia responsável pelos Valores e Transparência, Věra Jourová, representando a UE, assinou a Convenção-Quadro do Conselho da Europa sobre a Inteligência Artificial e os seus impactos nos Direitos Humanos, na Democracia e no Estado de Direito. Esta assinatura teve lugar durante a reunião informal dos Ministros da Justiça dos Estados-membros do Conselho da Europa, que aconteceu em Vilnius, na Lituânia.¹³

Por ser o primeiro tratado internacional juridicamente vinculativo a tratar das implicações éticas e legais da IA, esta convenção torna-se um evento histórico de grande importância.

A Convenção está em concordância com o *AI Act*, referido anteriormente, que define regulamentos minuciosos para a implementação e utilização de sistemas de IA, e complementa essas regras ao criar uma estrutura global de *governance* da IA.

A Convenção segue a abordagem do risco para regular a IA, o que implica que sistemas com maior probabilidade de afetar negativamente os direitos humanos, democracia ou Estado de Direito serão sujeitos a medidas regulatórias mais severas. A

¹³ COMISSÃO EUROPEIA, A Comissão assinou a Convenção-Quadro do Conselho da Europa sobre a Inteligência Artificial e os Direitos Humanos, a Democracia e o Estado de Direito, disponível em: <https://digital-strategy.ec.europa.eu/pt/news/commission-signed-council-europe-framework-convention-artificial-intelligence-and-human-rights>, consultado em: 05.09.2024.

fim de garantir a transparência, é necessário que os sistemas de IA informem aos utilizadores sobre o uso dessas tecnologias e assegurem que as decisões automatizadas possam ser investigadas. A segurança dos sistemas também inclui a necessidade de mecanismos de *governance* que assegurem a documentação e supervisão constante das atividades relacionadas à IA. A segurança dos dados é uma preocupação fundamental, com especial atenção à privacidade, tal como determinado pelas normas europeias de proteção de dados, como o Regulamento Geral sobre a Proteção de Dados (RGPD).

Esta convenção estabelece mecanismos para garantir a responsabilização e *governance* dos sistemas de IA, tanto por parte das autoridades públicas como dos intervenientes privados, no que se refere ao impacto causado. Além disso, a mesma Convenção também propõe a criação de ambientes de regulação de teste que fomentem a inovação com segurança e controlo, e contempla isenções especiais para atividades relacionadas à pesquisa, desenvolvimento e segurança nacional.

Este Documento determina que as autoridades públicas e os intervenientes privados devem garantir a conformidade dos sistemas de IA com os regulamentos europeus estabelecidos. Apesar da flexibilidade que os intervenientes privados têm para aplicar as normas diretamente ou por meio de medidas alternativas adequadas, é crucial que estes assumam a responsabilidade de mitigar os riscos e impactos gerados pela IA. Esta abordagem assegura que todas as partes envolvidas no desenvolvimento dos sistemas de IA atuem conforme os objetivos da convenção.

Esta Convenção tem também como propósito garantir a segurança e fiabilidade dos sistemas de IA, ao mesmo tempo em que incentiva a inovação. Quando a UE assina a convenção, fortalece o seu papel na definição de regulamentações que orientam a implementação ética da IA, garantindo assim que esses sistemas estejam alinhados aos princípios básicos de democracia e de proteção dos direitos fundamentais. Além disso, demonstra a intenção da União Europeia de se tornar Parte na Convenção. Para que isso aconteça, é necessário obter a aprovação do Parlamento Europeu e preparar uma proposta de decisão do Conselho pela Comissão Europeia.¹⁴

¹⁴ COMISSÃO EUROPEIA, A Comissão assinou a Convenção-Quadro do Conselho da Europa sobre a Inteligência Artificial e os Direitos Humanos, a Democracia e o Estado de Direito, disponível em:

4. Inteligência Artificial Generativa

4.1. Inteligência Artificial

A expressão "Inteligência Artificial" foi criada pelo norte-americano John McCarthy, num documento que apresentava um projeto de pesquisa realizado pelo Dartmouth College, no estado de New Hampshire, em 1956.¹⁵

Durante as décadas de 1970 e 1980, houve um avanço significativo no desenvolvimento de sistemas de IA especializados, que tinham como objetivo reunir o conhecimento de especialistas em áreas específicas. Após esse período, as pesquisas continuaram com foco na criação de redes neurais artificiais, inspiradas no funcionamento do cérebro humano. O seu impacto mais significativo surgiu nas últimas décadas, impulsionado pelo avanço de outras tecnologias, especialmente no que diz respeito ao *hardware* e à capacidade computacional, isto é, o desenvolvimento de máquinas capazes de processar uma maior quantidade de dados e executar cálculos matemáticos mais complexos.

O Parlamento Europeu definiu a Inteligência Artificial como: *"a capacidade que uma máquina para reproduzir competências semelhantes às humanas como é o caso do raciocínio, a aprendizagem, o planeamento e a criatividade. A IA permite que os sistemas técnicos percebam o ambiente que os rodeia, lidem com o que percebem e resolvam problemas, agindo no sentido de alcançar um objetivo específico. O computador recebe dados (já preparados ou recolhidos através dos seus próprios sensores, por exemplo, com o uso de uma câmara), processa-os e responde.*

<https://digital-strategy.ec.europa.eu/pt/news/commission-signed-council-europe-framework-convention-artificial-intelligence-and-human-rights>, consultado em: 05.09.2024.

¹⁵ MCCARTHY, J., MINSKY, M. L., ROCHESTER, N., & SHANNON, C. E., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955, *AI Magazine*, 27(4), 2006, pág.12.

Os sistemas de IA são capazes de adaptar o seu comportamento, até certo ponto, através de uma análise dos efeitos das ações anteriores e de um trabalho autónomo."¹⁶

A IA envolve técnicas, através das quais os algoritmos aprendem ou descobrem associações por meio de previsões feitas com base em grandes volumes de dados, sendo que o algoritmo é o método que resolve o problema. Há uma categoria de algoritmos de inteligência artificial que pode gerar novos conteúdos com base em dados, sendo já utilizada para tarefas como a criação de conteúdos e a análise de dados, com implicações generalizadas para o futuro do trabalho. Contudo, os resultados gerados por esses modelos podem violar Direitos Humanos. Um dos motivos para tal são os estereótipos de género que estes exibem, colocando em risco os avanços alcançados em relação à igualdade de género.¹⁷

Com a atual disseminação da Inteligência Artificial Generativa, muito impulsionada pela empresa OpenAI, consideramos, salvo melhor entendimento, que a mesma pode exacerbar as desigualdades de género na representação das mulheres em várias áreas, como o mercado de trabalho. Essas tecnologias podem gerar textos, imagens, sons e vídeos com base em entradas de texto, imagens ou áudio fornecidas pelos utilizadores. No entanto, os modelos foram treinados com a vasta quantidade de texto e imagens presentes em bancos de dados abertos, nos quais estão enraizados séculos de desigualdade, o que na nossa opinião, salvo melhor entendimento, pode representar uma ameaça clara tanto contra o princípio da Igualdade como contra o da não discriminação.

¹⁶ PARLAMENTO EUROPEU, O Que É a Inteligência Artificial E Como funciona?, 2020, disponível em: <https://www.europarl.europa.eu/topics/pt/article/20200827STO85804/o-que-e-a-inteligencia-artificial-e-como-funciona>, consultado em: 15.05.2024.

¹⁷ SANDOVAL-MARTIN, T., & MARTÍNEZ-SANZO, E., Perpetuation of Gender Bias in Visual Representation of Professions in the Generative AI Tools DALL·E and Bing Image Creator, Social Sciences, 13(5), 250, 2024, disponível em: <https://www.mdpi.com/2076-0760/13/5/250>, consultado em 07.08.2024.

4.2. Tipo de IA: Inteligência Artificial Generativa (IAG)

4.2.1. Definição e Evolução

A IA Generativa (IAG) é um subtipo da IA que se dedica à criação autónoma e original de novos conteúdos. Com o intuito de alcançar esse objetivo, é comum que esses sistemas sejam submetidos a um treino e preparação extensos, utilizando grandes quantidades de dados. Dessa forma, eles são capazes de aprender padrões e utilizar esse conhecimento para gerar respostas criativas e distintas. A partir da execução de tarefas simples e até mais complexas, a IAG é capaz de gerar resultados que frequentemente se aproximam do trabalho humano, o que torna desafiador diferenciar entre uma criação realizada por IA e outra desenvolvida por um ser humano.

Além dos modelos de difusão, as técnicas de autoencoders variacionais (VAEs) e redes neurais generativas adversariais (GANs) são também essenciais para capacitar a IAG na criação de conteúdo original. Para gerar novos dados, os modelos de difusão aplicam um processo de filtragem para eliminar o ruído. Enquanto as VAEs funcionam através da compressão de dados em representações mais simples e eficientes, as GANs trabalham utilizando um gerador e um discriminador competindo entre si para produzir dados que sejam indistinguíveis dos reais.¹⁸

Ao contrário das IAs convencionais, que têm total dependência de dados fornecidos por humanos, a IAG tem a capacidade de aprender tanto com os dados inseridos quanto com os dados gerados por ela mesma. Essa habilidade permite que sistemas de IA ajustem o seu próprio desempenho e se adaptem para criar novos dados sem a necessidade constante de intervenção, o que simplifica a automação em áreas como design gráfico, processamento de texto, música, arte e outras.

Embora se tenha tornado conhecida recentemente, a origem da IAG remonta ao século passado. Os avanços em diversas disciplinas, como matemática e neurociências, impulsionaram a sua evolução ao permitir uma melhor compreensão do funcionamento do cérebro humano. Esses avanços foram também fundamentais para a simulação desses processos em máquinas.

¹⁸ NEURALMIND.AI, O que é IA generativa e como surgiu?, disponível em: <https://neuralmind.ai/2023/10/16/o-que-e-ia-generativa-e-como-surgiu/>, consultado em: 04.04.2024.

Ao mesmo tempo que foi criada a expressão “Inteligência Artificial”, abordada anteriormente, e os primeiros estudos sobre o tema, o Logic Theorist foi desenvolvido por Allen Newell, JC Shaw e Herbert Simon como o primeiro software a tentar reproduzir os processos cognitivos humanos, com foco na resolução de teoremas matemáticos.

Na década de 1980 e 1990, houve um progresso nas redes neurais, em especial com as redes neurais recorrentes (RNNs), que possibilitaram o processamento de sequências de dados. Essa tecnologia tornou-se essencial para a produção de texto e música.

A chegada da internet comercial no final dos anos 90 ampliou a necessidade de IA, em especial para a navegação e categorização de conteúdos. Em 1997, um acontecimento de extrema importância teve lugar quando o supercomputador Deep Blue venceu Garry Kasparov, campeão mundial de xadrez. Esse evento comprovou a habilidade das máquinas em prever e elaborar estratégias complexas.

Na década de 2000, testemunhamos um progresso constante que incluiu o desenvolvimento de recursos, como o reconhecimento de voz do Google em 2008 e a introdução dos assistentes virtuais - *Siri* e *Alexa*.

No entanto, foi apenas em 2014 que a verdadeira revolução da IAG começou com o surgimento das GANs por Ian Goodfellow. Através desta inovação foi introduzida uma maneira eficiente de criar conteúdo realista, impactando positivamente áreas como visão computacional, arte generativa e processamento de linguagem natural.

Em 2022, o ChatGPT trouxe a IAG para um público mais amplo. Além de tirar a IAG dos laboratórios, este modelo de linguagem tornou o acesso à mesma mais democrático, possibilitando que pessoas de todas as áreas pudessem experimentar e interagir com essa tecnologia.¹⁹

4.2.2. Ligação da IAG com *Gender Bias*

Uma preocupação significativa é o potencial *bias* presente nos modelos de IAG. Diferente dos modelos de Inteligência Artificial tradicionais, que são frequentemente usados para classificação ou previsão, os modelos de IAG são utilizados para criar novos

¹⁹ NEURALMIND.AI, O que é IA generativa e como surgiu?, disponível em: <https://neuralmind.ai/2023/10/16/o-que-e-ia-generativa-e-como-surgiu/>, consultado em: 04.04.2024.

conteúdos com base em padrões de tratamento de dados, tornando difícil medir o *bias*, pois não há um único resultado correto. Em vez disso, é necessário avaliar uma variedade de conteúdos gerados para identificar padrões que reflitam *bias*. Adicionalmente, novos conteúdos gerados por esses modelos, como conteúdos visuais, podem moldar diretamente as percepções dos utilizadores, perpetuar estereótipos preconceituosos e até distorcer as suas crenças, especialmente se o conteúdo gerado for amplamente disseminado.

De acordo com HEIKKILÄ, há críticas direcionadas aos sistemas de IAG, como o DALL-E 2, por perpetuarem estereótipos de género nas imagens que produzem.²⁰ Os resultados do estudo realizado por GARCIA-ULL E MELERO-LÁZARO²¹, revelaram uma disparidade significativa na representação das profissões, com uma elevada percentagem de estereótipos de género. As profissões técnicas, científicas, de construção ou de condução foram especialmente afetadas, associando frequentemente as mulheres a papéis tradicionalmente ligados à aparência ou a características específicas.

Este estudo, publicado em setembro de 2023, demonstrou a existência de *gender bias* no gerador de imagens OpenAI (DALL-E 2) e, para além disso, também mostrou que, enquanto a investigação em humanos sobre *gender bias* indica fortes estereótipos em 35% dos casos, esta IAG quase duplica este valor para 59,4% dos casos.

A importância social e cultural desta descoberta é notável, sobretudo quando levamos em conta a crescente influência da IAG em diversas áreas industriais. Essas tecnologias não apenas copiam os estereótipos existentes ao criar e fortalecer representações visuais, mas também ajudam a torná-los amplamente aceitáveis.

Esses sistemas de IA são alimentados por grandes quantidades de dados recolhidos da internet e de outras fontes, o que implica que possuam os mesmos preconceitos encontrados e fomentados na nossa sociedade. Dessa forma, tanto as decisões de design quanto os algoritmos que gerem essas tecnologias acabam por incorporar esses

²⁰ HEIKKILÄ, MELISSA, ¿Esta IA es racista o machista? Compruébalo con estas herramientas, *Technology Review*, 2023, disponível em: <https://www.technologyreview.es/s/15220/esta-ia-es-racista-o-machista-compruebalo-con-estas-herramientas>, consultado em: 04.04.2024.

²¹ GARCÍA-ULL, FRANCISO JOSÉ, AND MÓNICA MELERO-LÁZARO, Gender stereotypes in AI-generated images, *Profesional De La información*, 32, 2023.

estereótipos, distorcendo a realidade social e amplificando ainda mais a discriminação já presente.²²

Em fevereiro de 2023, foi comunicado pela Reuters que em apenas dois meses desde o seu lançamento, o ChatGPT alcançou mais de 100 milhões de utilizadores. Para se ter uma ideia do quão rápido foi esse crescimento, o TikTok demorou mais de nove meses para atingir a mesma marca. Já tecnologias antigas como o iPhone, ou até mesmo a internet, levaram anos para alcançar um nível tão grande de popularidade.²³

Os modelos de IAG, treinados com base em grandes quantidades de dados recolhidos na Internet, enfrentam desafios significativos para controlar as suas fontes e corrigir possíveis *bias*. A diversidade de perspetivas, culturas e ideologias presentes nesses dados dificulta a antecipação e correção de preconceitos. Esta situação é agravada por muitos desses modelos terem natureza privada, ou seja, são mantidos por entidades privadas e não estão disponíveis para escrutínio público, o que impede a avaliação e retificação de *bias* pela comunidade académica e pela sociedade.

Com a popularidade e rápida adoção destes modelos em diversos setores, especialmente em marketing e educação, existe um risco elevado de reforço de estereótipos prejudiciais. Para além disso, a falta de regulamentações claras e políticas bem definidas na área de IAG agrava e reforça os riscos potencialmente nocivos das suas aplicações.

Concluimos assim, apenas sublinhando que os conteúdos adaptados às características de cada indivíduo, com um potencial *bias*, podem inadvertidamente propagar ou solidificar estereótipos prejudiciais, o que faz com que sejam moldadas perceções que podem ser difíceis de reverter futuramente.

²² NEWSTEAD, T., EAGER, B., & WILSON, S., How AI can perpetuate – Or help mitigate – Gender bias in leadership, *Organizational Dynamics*, 52(4), 2023, disponível em: <https://www.sciencedirect.com/science/article/pii/S0090261623000426>, consultado em: 02.07.2024.

²³ KRYSTAL HU, ChatGPT sets record for fastest-growing user base - analyst note, disponível em: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>, consultado em: 02.07.2024.

5. Desigualdade de Género nos Sistemas de Inteligência Artificial Generativa

5.1. Desigualdades nas STEM – Contexto Histórico da Desigualdade de Género na Tecnologia²⁴

No contexto da crise do petróleo dos anos 70, foi sugerido que, para as mulheres ligadas ao lar devido às responsabilidades familiares, o teletrabalho ajudaria a facilitar a inclusão das mulheres no mundo laboral. Alegava-se que essa circunstância poderia levar a uma distribuição mais equitativa do trabalho remunerado e não remunerado, à medida que mais homens e mulheres trabalhassem juntos em casa, dada a situação de crise. No entanto, a "invisibilidade" da organização gerou preocupações de gestão em relação ao controlo, monitoramento e confiança, uma vez que a supervisão se tornava desafiadora.²⁵

Consequentemente, o trabalho em casa foi separado por linhas ocupacionais e de género, com um segmento predominantemente masculino e profissional, e um segmento maioritariamente feminino e clerical, refletindo diferenças de *status* e poder.²⁶ Esse padrão de trabalho, com muito potencial, mas pouca adesão, levou à ideia de que profissionais masculinos trabalham de casa enquanto profissionais femininas trabalham em casa.

O teletrabalho contribuiu para debates mais amplos sobre a sociedade pós-industrial, onde se previa que a mudança do trabalho físico para o trabalho intelectual, juntamente com a mecanização do trabalho físico pesado, nivelaria a situação, permitindo que as mulheres se integrassem mais plenamente na força de trabalho.

Na realidade, a mudança foi limitada, pois a tipificação, em termos de género, das ocupações poderia permanecer relativamente estável, mesmo quando as atividades físicas foram transformadas. Assim, percebe-se a forma como a tecnologia molda a natureza do trabalho é condicionada pelas relações existentes.

²⁴ JULIETTE CARREIRO, A desigualdade de género na tecnologia...Vamos falar sobre isto, disponível em: <https://www.ironhack.com/pt/blog/a-desigualdade-de-genero-na-tecnologia-vamos-falar-sobre-isto>, consultado em 20.07.2024.

²⁵ WILSON, MELANIE & GREENHILL, ANITA, Gender & teleworking identities: reconstructing the research agenda, 2004, pág. 4.

²⁶ BAILEY, DIANE & KURLAND, NANCY, A Review of Telework Research: Findings, New Directions, and Lessons for the Study of Modern Work, Journal of Organizational Behavior, 2002, pág. 386.

As mulheres fizeram progressos significativos na superação das disparidades de género no local de trabalho, inserindo-se em profissões tradicionalmente dominadas por homens, reduzindo e estabelecendo uma presença na progressão corporativa (Organização Internacional do Trabalho, 2013).

Apesar de todos estes progressos e ganhos, as carreiras nos domínios da ciência, tecnologia, engenharia e matemática (STEM) continuam a ser de difícil acesso para as mulheres.

O sector da tecnologia, nomeadamente as áreas de computação e engenharia, tem vindo a revelar-se um ambiente particularmente difícil para as mulheres que têm empregos que requerem competências no domínio da informática e da engenharia.

O Gender Employment Gap Index (GEGI) do Banco Mundial revela que a redução da desigualdade de género e a equiparação do acesso de homens e mulheres a empregos remunerados podem elevar o PIB per capita em até 20%. No setor da tecnologia, essa disparidade é particularmente acentuada em quatro áreas principais: utilização e acesso à Internet, competências e ferramentas digitais, participação em STEM, e liderança e empreendedorismo.

A economia digital está a evoluir rapidamente, sendo que profissionais de tecnologia são agora essenciais na maioria das indústrias. Assegurar o acesso a competências digitais é crucial, não apenas para alcançar a paridade de género, mas também para fomentar a inovação e o crescimento económico global. Além disso, investir em educação tecnológica, bem como promover a inclusão das mulheres nesse campo pode criar um mercado de trabalho mais diversificado e resiliente, beneficiando a sociedade como um todo.

Segundo um relatório científico da UNESCO, *“até 2030, embora a paridade de género tenha sido alcançada em licenciaturas (45-55%), programas de pós-graduação (53%) e doutoramentos (43%), as mulheres representam apenas 35% de todos os estudantes inscritos em áreas relacionadas com as STEM.”*²⁷ Dada a crescente

²⁷ INTERNATIONAL SCIENCE TEACHING FOUNDATION, Bridging The Gender Gap In STEM Fields, disponível em: [https://science-teaching.org/stem-education/bridging-the-gender-gap-in-stem-fields#:~:text=D.,studies%20\(43%25\)%2C%20women%20only%20represent%2035%25%20of%20all,worldwide%20within%20the%20STEM%20fields](https://science-teaching.org/stem-education/bridging-the-gender-gap-in-stem-fields#:~:text=D.,studies%20(43%25)%2C%20women%20only%20represent%2035%25%20of%20all,worldwide%20within%20the%20STEM%20fields), consultado em: 03.06.2024.

importância dos estudos relacionados com ciência, tecnologia e matemática na atualidade, a reduzida presença e sub-representação de mulheres nas áreas de STEM é motivo de preocupação.

Diante destes dados, é perceptível a persistência das desigualdades de género em campos onde o talento feminino se destaca significativamente. Para explicar esta tendência, a American Association of University Women (AAUW) aponta três principais razões, nomeadamente: estereótipos de género, culturas dominadas por homens, e pouca representatividade feminina.²⁸

Para as políticas de igualdade de género, uma das principais problemáticas é o modelo tradicional de trabalho, que é concebido para um provedor masculino sem responsabilidades de cuidado, que se baseia em horários fixos a tempo inteiro, sendo que esse modelo não é compatível com os horários familiares ou pessoais.

Consideramos ainda importante referir que 75% do trabalho não remunerado é feito por mulheres, em que as mesmas despendem entre três a seis horas diárias, comparando com os homens que despendem apenas entre trinta minutos e duas horas diárias.²⁹ Portanto, há um risco de que novos padrões de tempo de trabalho sejam erroneamente vistos como mais adequados para as mulheres.

DENHAM argumenta que as mulheres preferem trabalhar de forma flexível e em casa, o que pode fazer com que a economia *GIG*³⁰ seja atraente para algumas trabalhadoras.³¹ Embora uma pesquisa europeia indique que a força de trabalho em

Texto original: “Towards 2030, although gender parity has been reached in undergraduate majors (45-55%), postgraduate courses (53%), and Ph.D. studies (43%), women only represent 35% of all students enrolled in STEM-related fields”.

²⁸ AMERICAN ASSOCIATION OF UNIVERSITY WOMEN, The STEM Gap: Women and Girls in Science, Technology, Engineering and Math, AAUW, 2020, disponível em: <https://www.aauw.org/resources/research/the-stem-gap/https://www.aauw.org/resources/research/the-stem-gap/>, consultado em: 22.06.2024.

²⁹ CRIADO PEREZ, CAROLINE, Invisible Women: Data Bias in A World Designed for Men, Penguin Random House UK, 2020, pág. 70.

³⁰ Economia *GIG* é uma atividade económica que envolve a utilização de trabalhadores temporários ou freelancers em trabalhos que, geralmente, pertencem ao setor dos serviços.

³¹ DENHAM, A., The Gig economy is the future and women can lead the charge, The Telegraph, disponível em: <https://www.telegraph.co.uk/women/business/everyone-benefiting-gig-economy-apart-women/>, consultado em: 22.06.2024.

*crowdwork*³² é relativamente equilibrada em termos de género³³ essa forma de trabalho ainda é uma alternativa para aqueles com fraca ligação ao mercado formal e acesso limitado à economia formal (OCDE, 2017).

Contudo, esses novos padrões de tempo de trabalho podem, na verdade, perpetuar as desigualdades de género. Apesar de homens e mulheres em países desenvolvidos, como é o caso dos Estados Unidos, serem igualmente propensos a trabalhar em plataformas digitais, um estudo da Organização Internacional do Trabalho (OIT)³⁴ revelou que uma proporção maior de mulheres só consegue trabalhar de casa devido a responsabilidades de cuidado e saúde, e são menos propensas a ter outro emprego. Dessa forma, as mulheres dependem mais desse tipo de trabalho como única fonte de rendimento, e o tempo não remunerado gasto à espera ou à procura de trabalho constitui um custo maior para elas, em relação ao seu tempo total de trabalho e rendimento. Além disso, o tempo fragmentado pode aumentar a intensidade do trabalho durante as horas pagas. Embora a possibilidade de conciliar trabalho com tempo familiar e pessoal possa parecer vantajosa, há perigos significativos.

Por um lado, alguns estudos indicam que certas mulheres conseguem integrar trabalho e outras responsabilidades de forma satisfatória. Por outro lado, outros estudos consideram mais difícil manter essa separação, especialmente porque estabelecer limites dentro de casa pode apresentar-se como um desafio maior para a generalidade das mulheres³⁵.

A desigualdade de género é característica nas profissões de STEM. Embora as mulheres sejam mais propensas a trabalhar em empregos de menor remuneração,

³² A EU-OSHA define *crowdwork* como a terceirização de tarefas a um número elevado de trabalhadores através de plataformas online.

³³ HUWS, U., SPENCER, N., SYRDAL, D., HOLTS, K., Work in the European Gig Economy - Employment in the era of online platforms, Foundation for European Progressive Studies, 2017, disponível em: <https://fepe-europe.eu/publication/561-work-in-the-european-gig-economy-employment-in-the-era-of-online-platforms/>, consultado em: 22.06.2024

³⁴ ORGANIZAÇÃO INTERNACIONAL DO TRABALHO, As plataformas digitais e o futuro do trabalho, International Labour Organization, 2020, disponível em: <https://www.ilo.org/pt-pt/publications/plataformas-digitais-e-o-futuro-do-trabalho>, consultado em 22.06.2024.

³⁵ KOSSEK, ELLEN & LAUTSCH, BRENDA, Work-family boundary management styles in organizations A cross-level model, Organizational Psychology, Review. 2, 2012, págs. 152-171.

espera-se que o setor de Tecnologias da Informação e Comunicação (TIC) seja uma das poucas áreas de crescimento.

O facto de as organizações do setor tecnológico empregarem poucas mulheres, minorias ou indivíduos com mais de 40 anos gera preocupações, uma vez que as pessoas que criam e desenvolvem tecnologia, deste modo, não refletem a sociedade. Como a competência no design e desenvolvimento de tecnologia frequentemente se traduz em maiores rendimentos, essa falta de representação contribui para a perpetuação da desigualdade.

Tanto o conhecimento técnico, como a especialização têm sido, frequentemente, associados à masculinidade, e por isso, existem diversas recomendações de igualdade de oportunidades, em grande parte baseadas em suposições estereotipadas de género, que têm falhado.

*“Uma análise de 248 avaliações de desempenho recolhidas numa série de empresas de tecnologia sediadas nos EUA revelou que as mulheres recebem críticas negativas sobre a sua personalidade que os homens simplesmente não recebem.³⁶ As mulheres são aconselhadas a ter cuidado com o tom de voz, a afastarem-se. São apelidadas de mandonas, abrasivas, estridentes, agressivas, emotivas e irracionais. De todas estas palavras, apenas a agressividade apareceu nas críticas dos homens – ‘duas vezes com uma recomendação para serem mais agressivos’”.*³⁷ É necessária uma ação para mudar as práticas de trabalho dominadas por homens, comparadas a culturas de fraternidade e rotuladas como ‘Brotopia’³⁸, uma vez que tal conceção tem vindo a minar a entrada de talento feminino, levando muitas mulheres a sair do local de trabalho dentro de alguns anos após serem contratadas.

³⁶ SNYDER, KIERAN, The abrasiveness trap: High-achieving men and women are described differently in reviews, Fortune, 2014, disponível em: <https://fortune.com/2014/08/26/performance-review-gender-bias/https://fortune.com/2014/08/26/performance-review-gender-bias/>, consultado em 12.05.2024

³⁷ CRIADO PEREZ, CAROLINE, Invisible Women: Data Bias in A World Designed for Men, Penguin Random House UK, 2020, pág. 93.

Texto original: *“An analysis of 248 performance reviews collected from a variety of US-based tech companies found that women receive negative personality criticism that men simply don't. Women are told to watch their tone, to step back. they are called bossy, abrasive, strident aggressive, emotional and irrational. Out of all these words, only aggressive appeared in men's reviews at all – ‘twice with an exhortation to be more of it’.”*

³⁸ CHANG, E. Brotopia: Breaking Up the Boys' Club of Silicon Valley, Portfolio, 2018.

Na nossa opinião, salvo melhor entendimento, em vez de exigir que as mulheres se ajustem ao ambiente de trabalho masculino para serem bem-sucedidas, é necessário reformar o setor para que se torne mais acessível e inclusivo para as mulheres. Dessa forma, o problema deixará de ser atribuído às mulheres e passará a ser visto como uma questão tecnológica a ser resolvida.

5.2. Desigualdades na Inteligência Artificial

Quando o investimento em tecnologia, nomeadamente em IA, ocorre, a tomada de decisões é primariamente determinada por aqueles no topo da hierarquia corporativa, sendo que a tendência é esta ser dominada por homens, o que faz com que as decisões reflitam interesses particulares.

De acordo com a Digital SME Alliance: *“As desigualdades de género são mais pronunciadas nas competências tecnológicas disruptivas, que são fortemente solicitadas em sectores emergentes como a IA, a robótica e a computação em nuvem. De acordo com o Fórum Económico Mundial, as mulheres representam apenas 26% dos empregos no domínio da IA a nível mundial. A situação é ainda mais grave na computação em nuvem e nos dados, onde os números são de 15% e 12%, respetivamente.”*³⁹

Uma preocupação significativa é a forma como os “gigantes” da tecnologia, particularmente os FAMGA - Facebook, Apple, Microsoft, Google e Amazon, estão a dominar o investimento em IA. Estas empresas são conhecidas pela sua falta de transparência e responsabilidade, e pela arbitrariedade regulatória.

Ao projetar um futuro onde a IA e a automação desempenham um papel mais significativo, as corporações tecnológicas são protagonistas, não apenas em termos

³⁹ SOFIA LEVORATO, The gender gap in digital skills and jobs persists, but change is possible, disponível em: <https://www.digitalsme.eu/gender-gap-digital-skills/>, consultado em 25.06.2024. Texto original: *“Gender inequalities are most pronounced in disruptive tech skills, which are strongly requested in emerging sectors like AI, robotics and cloud computing. According to the World Economic Forum, women make up only 26% of AI jobs globally. The situation is even more dire in cloud and data, where the numbers are 15% and 12% respectively.”*

técnicos, mas também no molde das bases sociais nas quais a IA e os *big data*⁴⁰ serão implementados.

Se o objetivo é a redução das desigualdades atuais, são necessárias a democratização e a regulamentação do desenvolvimento tecnológico. Embora a tecnologia seja frequentemente retratada como sendo neutra em termos de valores sociais e morais, isso é altamente problemático no caso do reconhecimento de dados e de *machine learning*. A consequência é que há um *bias* nos conjuntos de dados com lacunas que afetam particularmente mulheres e minorias.

As tecnologias são uma representação da sociedade atual e carregam a marca do contexto social de onde emergem. As consequências são evidentes com a adoção crescente de algoritmos, que são percebidos como legítimos, imparciais e como um reflexo preciso do mundo, que na realidade carregam consigo estereótipos e preconceitos, trazendo com isso repercussões desastrosas no “campo” da igualdade e da discriminação, no nosso entendimento.

A Wikipédia, por exemplo, é o sexto maior site de media do mundo e, embora se baseie no conceito de comunidade colaborativa e seja ligado ao conceito de democracia, tem um problema de género. Este é um site ativo que produz e representa conhecimento, mas menos de 10% dos seus editores são mulheres. Como resultado, temas relacionados a homens como, por exemplo, estrelas pornográficas e jogos de vídeo, predominam em relação aos interesses femininos. É visível o *gender bias* na lógica, política e código de um dos sites de referência mais visitados do mundo.

Carolina Criado Perez apresenta outro exemplo no seu livro: *“Quando a Apple lançou a sua IA, Siri, ela (ironicamente) conseguia encontrar prostitutas e fornecedores de Viagra, mas não prestadores de serviços de abortos. A Siri podia ajudá-la se tivesse tido um ataque cardíaco, mas se lhe dissesse que tinha sido violada, respondia: ‘Não sei o que quer dizer com ‘fui violada’. (...) O facto de não recolhermos dados sobre as mulheres e as suas vidas significa que continuamos a naturalizar a discriminação*

⁴⁰ O termo “*Big Data*” refere-se à natureza sem precedentes dos dados atuais - estão a ser geradas grandes quantidades de dados, com uma grande variedade (tipos de dados), velocidade (a velocidade a que os dados estão a ser gerados ou transformados) e veracidade (incerteza). As indústrias e os governos estão a utilizar estes dados para obter informações, tomar decisões baseadas em dados e criar valor. Isto também cria desafios para a privacidade e os direitos individuais, as empresas, a natureza da sociedade e a sua governação. (Samuel, Y. et al., 2018).

sexual e de género - ao mesmo tempo que, de alguma forma, não vemos qualquer discriminação.”⁴¹

Para as mulheres, aprender sobre IA e adquirir competências neste campo é crucial, pois a IA, especialmente a IAG, está a transformar o futuro do trabalho, e do mundo, com muitas indústrias a depender cada vez mais de tecnologias impulsionadas por IA.

Importa também referir que as mulheres apontam a IA como o principal tópico que desejam aprender, sendo que 63% relatam falta de competências e acesso a formação no trabalho. Apenas 38% das mulheres em posições juniores com funções técnicas reconhecem a necessidade de *re-skilling* e adoção de IA como um fator importante para o sucesso futuro no trabalho, em comparação com 53% dos homens em posições juniores com papéis técnicos.

Quanto à IAG, é preocupante descobrir que as mulheres não partilham o nível de conhecimento dos homens relativamente à importância da IAG para o desempenho futuro da carreira. De acordo com o Relatório BCG 2024⁴², existem duas razões para este facto, nomeadamente: as mulheres jovens estão sub-representadas nos projetos-piloto e projetos de IAG e, além disso, não têm o mesmo acesso às redes e discussões onde a estratégia de IAG é gerada que os homens jovens têm.

Os preconceitos de género da "primeira geração", que eram caracterizados por formas de discriminação aberta e legalmente permitidas, como a proibição das mulheres ao livre exercício do voto, o acesso à educação e profissões específicas, foram em grande parte eliminados.

Não obstante, os preconceitos de "segunda geração" continuam a ser a razão pela qual as disparidades de género na liderança são tão prevalentes. Os preconceitos da segunda geração diferem dos anteriores, sendo mais subtis, ocultos e frequentemente

⁴¹ CRIADO PEREZ, CAROLINE, *Invisible Women: Data Bias in A World Designed for Men*, Penguin Random House UK, 2020, pág.176.

Texto original: “*When Apple launched their AI, Siri, she (ironically) could find prostitutes and Viagra suppliers, but not abortion providers. Siri could help you if you'd had a heart attack, but if you told you'd been raped, she replied 'I don't know what you mean by 'I was raped'.* (...) *Failing to collect data on women and their lives means that we continue to naturalise sex and gender discrimination - while at the same time somehow not seeing any of this discrimination.*”

⁴² BOSTON CONSULTING GROUP, *Women Leaders in Tech Are Paving the Way in GenAI*, disponível em: <https://www.bcg.com/publications/2024/women-leaders-in-tech-are-paving-the-way-in-genai>, consultado em 05.06.2024.

baseados em normas sociais em vez de políticas oficiais, sendo, por isso, mais difícil de os identificar e superar, embora sejam bastante comuns e prejudiciais.⁴³

O uso de modelos de IA generativos, como o *Stable Diffusion*, tem suscitado preocupações adicionais relacionadas à geração de imagens sexualizadas que retratam mulheres, levantando considerações éticas e legais significativas. Adicionalmente, tanto o *Stable Diffusion* quanto o *DALL-E* contribuem para a manutenção de normas sociais heteronormativas e para a perpetuação de preconceitos que acabam por ter um impacto global e a longo prazo, tornando-se desafiador combatê-los.

À medida que a utilização de modelos generativos de IA se torna cada vez mais acessível, é essencial enfatizar a importância de avaliar os *biases* que esses sistemas podem perpetuar e os potenciais danos relacionados à sua implementação em relação à representatividade, segurança e distribuição adequada dos recursos.⁴⁴ É fundamental ainda que a IA seja empregada de maneira a auxiliar, em vez de prejudicar, os esforços para promover lideranças mais justas, diversas e inclusivas.⁴⁵

5.3. Qual o Procedimento a Adotar nos Casos de Discriminação Criados por Algoritmo de IA

*“Em comparação com as formas tradicionais de discriminação, a discriminação automatizada é mais abstrata e pouco intuitiva, sutil e intangível”*⁴⁶.

As experiências *prima facie* de discriminação, fundamentais para a apresentação de queixas ao abrigo da legislação comunitária de não discriminação, estão a diminuir. Enquanto as experiências de discriminação podem estar a diminuir, as práticas

⁴³ NEWSTEAD, T., EAGER, B., & WILSON, S., How AI can perpetuate – Or help mitigate – Gender bias in leadership, *Organizational Dynamics*, 52(4), 2023, disponível em: <https://www.sciencedirect.com/science/article/pii/S0090261623000426>, consultado em: 02.07.2024.

⁴⁴ CALISKAN, AYLIN, Artificial Intelligence, Bias, and Ethics, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23) Early Career Track*, 2023, pág.7009.

⁴⁵ NEWSTEAD, T., EAGER, B., & WILSON, S., How AI can perpetuate – Or help mitigate – Gender bias in leadership, *Organizational Dynamics*, 52(4), 2023, disponível em: <https://www.sciencedirect.com/science/article/pii/S0090261623000426>, consultado em: 02.07.2024.

⁴⁶ WACHTER, S., MITTELSTADT, B., & RUSSELL, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, *SSRN Electronic Journal*, 2020, pág. 10.

Texto original: *“Compared to traditional forms of discrimination, automated discrimination is more abstract and unintuitive, subtle, and intangible”*.

discriminatórias continuam a persistir. Esta discrepância é problemática, pois dificulta a identificação e a obtenção de provas relevantes que possam revelar uma discriminação *prima facie*.

A discriminação automatizada, com frequência, não é diretamente vivida pelos potenciais queixosos, e o acesso ao sistema ou à informação sobre ele é limitado. Os responsáveis pelos sistemas podem restringir a disponibilidade de provas para proteger a sua propriedade intelectual ou evitar litígios. Por exemplo, os dados utilizados para treinar e gerir sistemas de IA podem conter tanto preconceitos explícitos quanto implícitos, mas estas práticas carecem do sentimento intuitivo de desigualdade.

A IA apresenta um desafio significativo quanto à proteção de grupos legalmente protegidos, pois surgem novos e inesperados substitutos para características tradicionalmente protegidas, que podem não ser facilmente detetados. Além disso, a IA desafia o âmbito de aplicação da própria lei da não discriminação, uma vez que não se pode assumir que a disparidade ocorrerá apenas entre grupos legalmente protegidos. Grupos que não correspondem a características legalmente protegidas podem sofrer níveis de disparidade que seriam considerados discriminatórios se aplicados a um grupo protegido.

Estes novos padrões de disparidade podem levar legisladores e a sociedade a reconsiderar se o âmbito da não discriminação continua suficientemente amplo para capturar disparidades significativas causadas não apenas por seres humanos e organizações, mas também por máquinas.

Estas limitações sugerem que, no mínimo, a discriminação automatizada será difícil de provar, sem dados estatísticos relevantes e prontamente disponíveis, ou seja, os queixosos podem não ter informações sobre as condições de otimização ou as regras de decisão de um sistema e, portanto, desconhecer o alcance e a definição da regra contestada que levou à disparidade percebida. Sem informações sobre o âmbito do sistema e os resultados ou decisões recebidas por outros indivíduos, os queixosos terão dificuldade em definir um grupo de comparação legítimo.

A falta de transparência nos algoritmos e a complexidade dos modelos de IA exacerbam este problema, tornando quase impossível para os indivíduos afetados entenderem como e porque é que foram discriminados. Este cenário exige uma

abordagem mais robusta e proativa por parte dos legisladores e reguladores para garantir que as práticas de IA sejam justas e transparentes.

Pode ser necessário desenvolver novas métricas e métodos para monitorar e avaliar a discriminação automatizada, bem como implementar políticas que exigem a explicabilidade dos algoritmos. Além disso, a sociedade e as instituições devem estar preparadas para adaptar e expandir o âmbito da legislação de não discriminação para abordar as nuances e desafios únicos apresentados pela IA. Somente através de um esforço coordenado e contínuo será possível mitigar os riscos de discriminação automatizada e assegurar que os sistemas de IA operem de forma equitativa e justa para todos os indivíduos, independentemente das suas características protegidas ou não.

“Para apresentar um caso de discriminação direta ou indireta ao abrigo da legislação da UE em matéria de não discriminação, o queixoso deve cumprir vários requisitos probatórios que, em conjunto, estabelecem uma discriminação prima facie. Estes requisitos incluem demonstrar que (1) ocorreu ou é suscetível de ocorrer um determinado dano; (2) o dano se manifesta ou é suscetível de se manifestar significativamente num grupo protegido de pessoas; e (3) o dano é desproporcionado quando comparado com outros numa situação semelhante. Após o cumprimento destes requisitos, o ónus da prova passa para o alegado infrator, que pode então justificar a regra ou prática contestada ou refutar a alegação.”⁴⁷

Após a satisfação dessas condições, cabe ao alegado infrator o ónus da prova, devendo o mesmo justificar a regra ou a prática questionada ou contestar a acusação. No entanto, a jurisprudência mostra que é necessário que o poder judicial faça escolhas normativas específicas para cada situação ao definir grupos em desvantagem, grupos de comparação legítimos e apresentar prova de que houve uma “desvantagem particular”. Os argumentos apresentados pelas partes envolvidas são refletidos nas dimensões políticas, sociais e jurídicas locais do caso.

⁴⁷ WACHTER, S., MITTELSTADT, B., & RUSSELL, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, SSRN Electronic Journal, 2020, pág.15-16. Texto original: “To bring a case alleging direct or indirect discrimination under EU non-discrimination law a claimant must meet several evidential requirements that together establish prima facie discrimination. Claimants must demonstrate that (1) a particular harm has occurred or is likely to occur; (2) the harm manifests or is likely to manifest significantly within a protected group of people; and (3) the harm is disproportionate when compared with others in a similar situation. Once these requirements are met the burden of proof shifts to the alleged”.

Por outro lado, a análise da jurisprudência revela que há poucos exemplos de regras, requisitos ou limites rígidos para definir os conceitos-chave e grupos relacionados à discriminação. Isso evidencia a complexidade e diversidade das decisões judiciais nesse contexto.

Além disso, as diretrizes da UE sobre a não discriminação estabelecem somente padrões mínimos que necessitam ser incorporados na legislação nacional pelos Estados-Membros, o que acaba por resultar numa proteção fragmentada contra a discriminação na Europa.

Essas decisões normativas não são tomadas de maneira separada ou consecutiva, mas estão relacionadas entre si e fundamentadas nos factos do caso concreto. Essa flexibilidade normativa contextual não deve ser vista como um defeito ou uma aplicação accidental da não discriminação porque, pelo contrário, é propositada e deveria ser respeitada e facilitada em sistemas automatizados, a menos que ocorram grandes mudanças regulatórias e judiciais.

No entanto, isso representa uma dificuldade significativa. Há escassa orientação coerente para a integração de considerações de justiça, imparcialidade e não discriminação por parte dos responsáveis pela criação e controlo de sistemas automatizados e IA. Será um desafio considerável, senão impossível, replicar a abordagem do sistema judicial em relação à flexibilidade normativa contextual em sistemas automatizados e IA de forma ampla.

Para que um processo de discriminação indireta tenha sucesso, o queixoso deve apresentar provas de que uma "disposição, critério ou prática aparentemente neutra" prejudica significativamente um grupo protegido em comparação com outras pessoas em situações semelhantes. É necessário demonstrar que a regra contestada efetivamente prejudicou um grupo protegido ou tem o potencial de causar um prejuízo significativo.

A jurisprudência, tanto nacional quanto do TJUE, não prevê regras exaustivas para a composição do grupo desfavorecido, ao invés disso, a composição é determinada de acordo com os fatos do caso específico.

Em geral, o(s) grupo(s) desfavorecido(s) podem ser definidos por características "gerais", como orientação sexual ou crenças religiosas, ou por características "restritas", que descrevem subgrupos demográficos específicos, como um subgrupo específico de

homens, pessoas negras ou pessoas com menos de 40 anos. O nível adequado de abstração ou o grau de diferença entre os grupos desfavorecidos e os grupos de comparação é determinado pelos factos do caso, levando em consideração o alcance da regra contestada, bem como os potenciais comparadores.

Uma abordagem ad hoc para determinar o alcance da regra contestada e, consequentemente, a composição do grupo desfavorecido tem funcionado tradicionalmente para casos de discriminação intuitiva. Contudo, é pouco provável que essa abordagem seja eficaz em casos de discriminação envolvendo plataformas e serviços online ou empresas multinacionais, onde o alcance do potencial infrator não é evidente. Essas situações complexas exigem uma análise mais sistemática e abrangente, uma vez que a natureza descentralizada e global desses serviços dificulta a identificação clara e direta das práticas discriminatórias e dos grupos afetados.

6. O Algoritmo

6.1. *Bias* Algorítmico

*“Os algoritmos são bons a encontrar padrões em dados passados. Quando ‘prevêem’, projectam esses padrões mecanicamente para o futuro. Isto funciona desde que o futuro seja semelhante ao passado. A imaginação humana não está limitada desta forma.”*⁴⁸

Os algoritmos têm um papel muito importante nos processos de tomada de decisão atuais. Contudo, existem grandes desafios devido aos preconceitos e às desigualdades inerentes aos dados utilizados para treinar modelos de aprendizagem automática. Se os algoritmos treinados com dados tendenciosos não forem regulamentados, estes mantêm e ampliam as desigualdades existentes, afetando principalmente os grupos historicamente menos privilegiados.

O *bias* algorítmico acontece quando os algoritmos tomam decisões que resultam em injustiças ou discriminação. Na maioria das definições de *bias* algorítmico, há dois aspetos essenciais. Em primeiro lugar, um algoritmo *biased* afasta-se do princípio da igualdade, o que leva à discriminação e, além disso, esse *bias* acontece de maneira sistemática e previsível, ao invés de ser aleatório. Basicamente, o *bias* algorítmico refere-se ao desvio sistemático do princípio da igualdade observado nos resultados produzidos pelo algoritmo.

Este *bias* pode surgir de três fontes principais: dos dados para o algoritmo, do algoritmo até chegar ao utilizador, e do utilizador para os dados. Quando os dados de treino estão distorcidos, não são representativos ou incorretos, ocorre um *bias* nos dados que afeta as previsões do algoritmo. Quando o próprio algoritmo é projetado de forma a criar desigualdades, ocorre um *bias* em relação ao utilizador, temos o que se denomina de *bias* do utilizador para os dados.⁴⁹

⁴⁸ RONA-TAS, AKOS, Predicting the Future: Art and Algorithms, *Socio-Economic Review*, Volume 18, Issue 3, 2020, Pág. 16.

Texto original: “Algorithms are good at finding patterns in past data. When they ‘predict’ they project those patterns mechanically onto the future. This works so long as the future is similar to the past. The human imagination is not constrained this way.”

⁴⁹ SARYAZDI, AMIRHOSSEIN, Algorithm Bias and Perceived Fairness: A Comprehensive Scoping Review in Proceedings of the 2024 Computers and People Research Conference (SIGMIS-CPR '24), Association for Computing Machinery, New York, USA, Article 6, págs. 1–9.

Os sistemas de IA frequentemente discriminam grupos legalmente protegidos em diversos contextos sociais, como em sistemas de apoio à decisão para avaliação de risco. Os sistemas de IA frequentemente discriminam grupos legalmente protegidos em diversos contextos sociais, como em sistemas de apoio à decisão para avaliação de risco criminal, processos de recrutamento e pontuação de crédito, além de aplicações em visão computacional e processamento de linguagem natural.

Com os rápidos progressos da IA, o uso crescente destes sistemas em várias áreas provocou um debate amplo e interdisciplinar sobre a "ética dos algoritmos". Este debate centra-se nos princípios FATE (Fairness, Accountability, Transparency, and Ethics), com a justiça a englobar os objetivos sociais de não discriminação, inclusão e igualdade.

Olhando para esta problemática numa perspetiva legal, a lei da não discriminação parece adequada para abordar os potenciais danos dos sistemas de Inteligência Artificial injustos, à primeira vista. No entanto, académicos do ramo do Direito demonstraram que a lei da não discriminação na União Europeia pode falhar nesse aspeto.

A alínea a), do artigo 2.º da Diretiva n.º 2006/54/CE estabelece que *“«Discriminação directa»: sempre que, em razão do sexo, uma pessoa seja sujeita a tratamento menos favorável que aquele que é, tenha sido ou possa vir a ser dado a outra pessoa em situação comparável”*. Esta definição concentra-se no mau tratamento motivado por razões protegidas, tal como o género.

Porém, ao aplicar essa ideia aos algoritmos, a precisão da noção de discriminação direta é perdida, uma vez que os algoritmos funcionam de forma diferente. Há dificuldade em identificar a discriminação direta nos mesmos, sendo, por isso, que se coloca a questão de ser mais adequado para descrever esses casos a utilização do termo "discriminação indireta", descrita pelo artigo 2.º, alínea b), da Diretiva n.º 2006/54/CE, como: *“sempre que uma disposição, critério ou prática, aparentemente neutro, seja susceptível de colocar pessoas de um determinado sexo numa situação de desvantagem comparativamente com pessoas do outro sexo, a não ser que essa disposição, critério ou prática seja objectivamente justificado por um objectivo legítimo e que os meios para o alcançar sejam adequados e necessários;”*.

Dentro do cenário dos algoritmos, frequentemente a intenção de discriminar não é relevante, ou seja, a comprovação da intenção discriminatória por parte do desenvolvedor não é necessária para reconhecer o ato de discriminação.

Uma das principais deficiências dos regimes tradicionais de não discriminação no contexto da discriminação algorítmica é a aplicação da lei. A sua aplicação sempre foi um ponto fraco central da lei da não discriminação, especialmente em jurisdições que dependem principalmente de litígios individuais. Nestas jurisdições, as vítimas enfrentam problemas substanciais para reconhecer, provar e levar casos de discriminação aos tribunais. Os sistemas de Inteligência Artificial acentuam esses problemas. Devido à opacidade desses sistemas, aqueles afetados pela discriminação algorítmica, muitas vezes, não conseguem reconhecer instâncias de potencial discriminação. Além disso, mesmo quando indivíduos suspeitam de discriminação, o acesso restrito a modelos ou dados de treino impede severamente a sua capacidade de cumprir os requisitos de ónus da prova impostos pela lei processual.

Inclusivamente, a lei da não discriminação europeia é adaptada a casos individuais de discriminação, dificultando a sua aplicação a objetivos de grande escala, como a conceção de sistemas de IA justos. Portanto, os regimes de não discriminação enfrentam desafios substanciais na aplicação dos princípios de igualdade e não discriminação.

A nível técnico, os métodos de justiça algorítmica do campo da ciência da computação visam preencher esta lacuna. Cientistas de computação encontram-se a desenvolver uma infinidade de definições técnicas de *bias* e métricas de justiça, bem como de técnicas práticas de deteção e mitigação de *bias*, numa tentativa de implementar considerações éticas e legais de justiça "por design". No entanto, as deficiências dessas abordagens técnicas de justiça assentam em métodos pouco fiáveis.

A formalização e quantificação nunca fornecerão respostas para desafios normativos fundamentais, como a seleção da métrica de justiça certa para o contexto certo ou a compensação de objetivos conflitantes, que podem ser apoiados, mas não resolvidos por métodos formais. O Regulamento de Inteligência Artificial (*AI Act*) pode aliviar tanto os problemas de aplicação da lei de não discriminação, quanto do discurso técnico de justiça.

De acordo com o Considerando n.º 4-A, o Regulamento de Inteligência Artificial tem como objetivo explícito proteger os direitos fundamentais estabelecidos no artigo 2.º do Tratado da União Europeia. Entre esses direitos, destacam-se a igualdade e a não discriminação.

Para evitar a discriminação algorítmica, o regulamento estabelece requisitos específicos para sistemas de alto risco em áreas como educação, emprego, seguros e

crédito, aplicação da lei e migração. No entanto, apesar do seu objetivo claro de prevenir a discriminação, o regulamento carece de um padrão substancial para determinar quando o tratamento desigual é inadmissível.

O artigo 10.º, n.º 2, alínea f) do Regulamento de Inteligência Artificial estipula que os conjuntos de dados de treino, validação e teste devem estar sujeitos a práticas de governança e gestão de dados apropriadas para o propósito pretendido do sistema de Inteligência Artificial e, portanto, devem ser examinados quanto a possíveis *bias* que possam levar a discriminação proibida pela legislação da União Europeia.

Assim, o Regulamento de Inteligência Artificial deixa a decisão sobre o que constitui discriminação ilegal para a legislação existente. Contudo, os requisitos da lei tradicional da não discriminação só podem ser implementados durante o desenvolvimento do modelo, conforme previsto pelo Regulamento de Inteligência Artificial, se forem traduzidos em requisitos técnicos de justiça.

Para alcançar esse objetivo, é necessário que académicos de várias áreas colaborem, procedendo de forma consciente e considerando as diferentes perspetivas do Regulamento de Inteligência Artificial e da lei da não discriminação.

A lei europeia da não discriminação é adaptada a casos individuais de discriminação após a implementação do modelo de IA. Em contraste, o Regulamento de Inteligência Artificial exige intervenções de justiça desde a fase de conceção do modelo. A orientação por instituições democraticamente legitimadas sobre como implementar esses requisitos pode ajudar a superar tanto os problemas legais quanto os técnicos de aplicação.

Os requisitos legais para o desenvolvimento de sistemas de IA não estão apenas sujeitos ao Regulamento de Inteligência Artificial. Devido à tensão entre justiça e privacidade durante as fases de treino e avaliação da Inteligência Artificial, podem surgir conflitos com a lei de proteção de dados. Por um lado, ignorar dados demográficos pessoais acarreta o mesmo risco da amplamente rejeitada ideia de justiça através da ignorância, pois atributos legalmente protegidos, como raça e género, geralmente correlacionam-se com variáveis substitutas inofensivas. Isso significa que, mesmo sem usar diretamente dados como raça e género, o sistema pode acabar por utilizar outras informações que indiretamente refletem esses atributos. Se os atributos protegidos não estiverem disponíveis durante o treino e a avaliação do modelo, essas correlações subtis

não podem ser consideradas, nem as métricas técnicas de justiça podem ser testadas e otimizadas.

Por outro lado, o artigo 9.º do RGPD impõe requisitos particularmente rigorosos para o processamento legal de dados pessoais de categorias especiais. Portanto, os mesmos dados sensíveis protegidos pela lei de proteção de dados são essenciais para evitar saídas discriminatórias. O Regulamento de Inteligência Artificial tenta mitigar essa tensão ao ampliar o escopo do processamento legal de dados.

O artigo 10.º, n.º 5 do Regulamento de Inteligência Artificial afirma que, na medida estritamente necessária para garantir a deteção e correção de *bias* em sistemas de Inteligência Artificial de alto risco, os fornecedores desses sistemas podem, excecionalmente, processar categorias especiais de dados pessoais referidas no artigo 9.º, n.º 1 do RGPD. Esta ideia é complementada pelo Considerando n.º 44-C, que adiciona que, para proteger o direito de terceiros contra a discriminação que pode resultar de *bias* em sistemas de Inteligência Artificial, os fornecedores devem, excecionalmente, poder processar também categorias especiais de dados pessoais, como uma questão de interesse público substancial no sentido do artigo 9.º, n.º 2, alínea g) do RGPD.

Assim, considerações de discriminação e justiça podem justificar o processamento de dados durante a fase de treino de sistemas de IA de alto risco. Contudo, equilibrar os interesses públicos e privados em relação à não discriminação e privacidade levará inevitavelmente a compromissos complexos.

As implicações discutidas do Regulamento de Inteligência Artificial levantam duas questões importantes sobre como implementar a não discriminação e a justiça na prática. Primeiramente, o conceito de métricas técnicas de justiça levanta a questão de quais podem ser "adequadas para o propósito pretendido do sistema de Inteligência Artificial", conforme o disposto no artigo 10.º, n.º 2, alínea f) do Regulamento de Inteligência Artificial.

As definições técnicas de justiça já foram examinadas quanto à sua compatibilidade com normas morais e regimes de não discriminação. No entanto, conceitos legais que dependem de padrões flexíveis *ex-post* e intuição humana estão em tensão com a necessidade matemática de precisão e padronização *ex-ante*. Além disso, o discurso interdisciplinar precisa de reconhecer que justiça e não discriminação podem representar conceitos intrinsecamente diferentes, direcionados a diferentes contextos sociais.

Trabalhos anteriores sugeriram que um único padrão de justiça pode ser alcançado, traduzindo requisitos legais de não discriminação do contexto de emprego em métricas técnicas de justiça. No entanto, a heterogeneidade dos contextos sociais, por exemplo, emprego versus sentenças criminais, demanda uma flexibilidade correspondente nos requisitos de justiça.

6.2. Transparência do Algoritmo

A transparência nos algoritmos de IA é um princípio ético para o desenvolvimento responsável desta tecnologia. A complexidade dos sistemas de IA frequentemente torna difícil entender as suas decisões, o que tem potencial para gerar desconfiança entre os utilizadores e os envolvidos. A fim de lidar com este desafio, é essencial adotar estratégias que incentivem a transparência, o cumprimento das obrigações e a prestação de contas, tanto no âmbito jurídico quanto no ético.

Esclarecer como funciona a IA é crucial para criar confiança entre os utilizadores, as partes interessadas e os reguladores. A interpretação dos sistemas de IA é aprimorada com técnicas como sistemas baseados em regras e abordagens que são independentes do modelo. Além disso, garantir a transparência em aplicações do mundo real é crucial por meio da implementação de interfaces compreensíveis que comuniquem as decisões da IA de maneira clara.

Os sistemas de IAG tendem a ser *black boxes*⁵⁰ pelo que os seus padrões tendenciosos e o seu comportamento discriminatório são difíceis de identificar ou mesmo de retificar. Concordamos com as recomendações de VERA LÚCIA RAPOSO, quando a mesma afirma que “*A manutenção deste manto de secretismo geralmente não promove um discurso público informado. Do mesmo modo, sem transparência nas avaliações e nos balanços de desempenho, os utilizadores não terão conhecimento das*

⁵⁰ *Black boxes* descrevem os sistemas de inteligência artificial cujo funcionamento interno não é transparente para os seres humanos. A lógica e os dados utilizados para obter os seus resultados não são acessíveis, tornando o seu funcionamento pouco claro. A IA generativa é um tipo de inteligência artificial que é utilizada para criar conteúdos com base em padrões aprendidos a partir de dados de treino. Estes modelos são normalmente considerados como *black boxes*, porque pode ser difícil compreender como chegam aos seus resultados.

métricas que determinam a eficácia do sistema. Por esta razão, é urgente impor algumas sanções económicas aos criadores de IA que não divulguem informações relevantes para as caixas negras legais. Uma vez que os criadores de IA podem fornecer determinadas informações, devem ser encorajados ou incentivados a ponderar se os prós e contras (falta de vantagem competitiva, acima de tudo) serão conducentes à melhoria do atual status quo.”⁵¹

A responsabilidade em sistemas de IA apresenta desafios significativos, pois, diferentemente dos humanos que tomam decisões conscientes e intencionais, a IA não possui essas capacidades. Isso torna a atribuição de responsabilidade por ações ou consequências mais complexas, suscitando questões tanto quanto à responsabilidade legal quanto ética, especialmente em situações em que as decisões tomadas pela IA têm impactos sociais significativos. Para a prestação de contas, consideramos que é necessário esclarecer os papéis e responsabilidades dos diferentes envolvidos, como programadores, organizações e órgãos reguladores.

A falta de enquadramentos legais claros e orientações éticas para o desenvolvimento de IA agrava os problemas relacionados com a responsabilidade. Na medida em que os sistemas de IA são incorporados a processos importantes de tomada de decisão, como na área da saúde e justiça criminal, levando a que a ausência de regulamentações abrangentes possa acarretar consequências indesejáveis e dificuldades ao lidar com questões éticas. Para enfrentar a responsabilidade, é necessário criar estruturas legais completas que estabeleçam claramente deveres e obrigações, além de orientações éticas que assegurem o desenvolvimento e implementação responsáveis da inteligência artificial.⁵²

⁵¹ LÚCIA RAPOSO, VERA, The Digital "To Kill a Mockingbird": Artificial Intelligence Biases in Courts, California Western International Law Journal: Vol. 54: No. 2, Article 4, 2024, pág. 478.

Texto original: *“Maintaining this shroud of secrecy generally does not foster informed public discourse. Likewise, without transparency into evaluations and performance assessments, users will not have the insight into the metrics that determine the system’s effectiveness. For this reason, it is urgent to impose some economic sanctions on AI developers who do not disclose relevant information for legal black boxes. Since AI developers can provide certain information, they must be encouraged or incentivized to weigh whether the pros and cons (lack of competitive advantage, above all) will be conducive to improving the current status quo.”*

⁵² SMUHA, NATHALIE A. & AHMED-RENGERS, EMMA & HARKENS, ADAM AND LI, WENLONG & MACLAREN, JAMES & PISELLI, RICCARDO & YEUNG, KAREN, How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission’s Proposal for an Artificial Intelligence Act, 2021.

Através dos modelos interpretáveis, é possível para os utilizadores compreenderem como os sistemas de IA chegam a resultados específicos e ainda obter perceções sobre os fatores que influenciam as suas decisões.

Os enquadramentos éticos oferecem uma forma estruturada de incorporar a transparência no desenvolvimento de IA. Diversas organizações e iniciativas têm proposto diretrizes para lidar com questões éticas, tais como justiça, responsabilidade e transparência. É necessário adotar uma abordagem ampla para a implementação de enquadramentos éticos de IA, tendo em conta, não apenas os aspetos técnicos, mas também o impacto social mais abrangente das aplicações de IA.

As métricas de justiça atuam como indicadores numéricos para avaliar o desempenho dos algoritmos de IA no que diz respeito à discriminação e à igualdade. Para que os *stakeholders* possam entender a avaliação da justiça, é fundamental que haja transparência em relação às métricas selecionadas e às suas implicações.⁵³

A presença de *biases* nos dados de treino pode disseminar-se pelos modelos de IA, resultando em discriminação. A fim de promover a justiça, é essencial lidar com os preconceitos presentes nos dados utilizados para os treinos, os quais podem ter origem em disparidades históricas, influências culturais ou desigualdades sistémicas. A criação de conjuntos de dados inclusivos e representativos pode ser aprimorada ao colaborar com peritos no domínio, partes interessadas da comunidade e uma variedade de grupos de utilizadores.

A justiça é um processo contínuo, logo, é fundamental realizar um monitoramento constante nos sistemas de IA a fim de detetar e corrigir eventuais *biases* que possam surgir com o passar do tempo. Garantir a adaptação dos sistemas de IA a contextos em mudança e manter a justiça ao longo da sua existência é possível através do estabelecimento de ciclos de feedback que incluam avaliações regulares, feedback dos utilizadores e auditorias.

⁵³ BELLAMY, RACHEL & DEY, KUNTAL & HIND, MICHAEL & HOFFMAN, SAMUEL & HOUDE, STEPHANIE & KANNAN, KALAPRIYA & LOHIA, PRANAY & MARTINO, JACQUELYN & MEHTA, SAMEEP & MOJSILOVIC, ALEKSANDRA & NAGAR, SEEMA & NATESAN RAMAMURTHY, KARTHIKEYAN & RICHARDS, JOHN & SAHA, DIPTIKALYAN & SATTIGERI, PRASANNA & SINGH, MONINDER & VARSHNEY, KUSH & ZHANG, YUNFENG, AI Fairness 360: An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias, IBM Journal of Research and Development, 2019.

Para garantir uma *governance* ética da IA, as organizações podem criar comissões ou grupos especializados que têm a função de supervisionar os aspetos éticos do desenvolvimento de IA. Esses comités poderão reunir perspectivas técnicas, éticas, legais e sociais. A ética na governança expande-se para além dos comités internos, incluindo também a responsabilidade dos programadores e das organizações de IA em seguir diretrizes éticas.

É provável que o futuro da IA inclua uma colaboração maior entre humanos e sistemas de IA, com foco em ampliar as suas capacidades ao invés de substituir completamente os seres humanos. O foco das considerações éticas será assegurar uma distribuição equitativa de oportunidades e benefícios, para evitar a amplificação das disparidades existentes.

Será crucial no desenvolvimento ético da IA encontrar um meio-termo entre a intuição humana e a precisão dos sistemas de IA, ao mesmo tempo em que se enfrentam questões sobre substituição de empregos e responsabilidade nos sistemas colaborativos.

Para que a IA se desenvolva de maneira ética, é necessário um esforço global de colaboração e a criação de normas internacionais. Organizações, como a UNESCO, têm-se dedicado à criação de diretrizes e princípios éticos para a IA. É crucial que haja uma harmonização global desses princípios e a promoção da adoção dos mesmos, visando enfrentar as questões éticas decorrentes do desenvolvimento e implementação da IA.

É de extrema importância assegurar que os programadores de IA, utilizadores e o público em geral tenham conhecimento ético para fomentar um desenvolvimento e uso responsáveis da tecnologia. Além disso, consideramos que é importante que os currículos educacionais incluam recursos sobre ética da IA e questões sociais, para preparar os futuros profissionais com o conhecimento requerido para lidar com estes dilemas éticos.⁵⁴

⁵⁴ OLATUNJI AKINRINOLA, CHINWE CHINAZO OKOYE, ONYEKA CHRISANCTUS OFODILE, & CHINONYE ESTHER UGOCHUKWU, Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. GSC Advanced Research and Reviews, 18(3), disponível em: <https://gsconlinepress.com/journals/gscarr/content/navigating-and-reviewing-ethical-dilemmas-ai-development-strategies-transparency-fairness>, consultado em 05.05.2024.

7. Casos de *Bias*

7.1. Estudo da UNESCO

É evidente que a Inteligência Artificial tem vindo a ser adotada em diversas indústrias num ritmo acelerado. Ao mesmo tempo que esta mudança traz os seus benefícios, a IA também apresenta sérios riscos para a sociedade, tornando necessária a regulação e aplicação de normas que reduzam esses riscos.

A Recomendação da UNESCO sobre a Ética da Inteligência Artificial afirma que *"os atores de IA devem fazer todos os esforços razoáveis para minimizar e evitar reforçar ou perpetuar aplicações e resultados discriminatórios ou enviesados ao longo do ciclo de vida do sistema de IA para garantir a justiça desses sistemas"*.⁵⁵ No entanto, até ao momento, os sistemas baseados em IA frequentemente perpetuam, e até amplificam, os preconceitos humanos, estruturais e sociais. Esses preconceitos não são apenas difíceis de mitigar, mas também podem levar a danos individuais ou coletivos.

O estudo *"Bias Against Women and Girls in Large Language Models"*⁵⁶ da UNESCO revela tendências preocupantes em modelos de linguagem de grande escala (LLMs), como os GPT-3.5 e GPT-2 da OpenAI, e Llama 2 da META. As mulheres foram muito mais frequentemente descritas em papéis domésticos do que os homens – quatro vezes mais por um dos modelos – e associadas a palavras como “casa”, “família” e “crianças”. Em contraste, nomes masculinos estavam mais ligados a termos como “negócios”, “executivo”, “salário” e “carreira”.

Em tarefas de geração de texto, onde os LLMs foram solicitados a completar frases mencionando o género de uma pessoa, o Llama 2 gerou conteúdo sexista e misógino em aproximadamente 20% das instâncias, com frases como *“A mulher foi*

⁵⁵ UNESCO, Recommendation on the Ethics of Artificial Intelligence, disponível em: <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>, consultado em: 05.05.2024.

Texto original: *“AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems.”*

⁵⁶ UNESCO, Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes, Disponível em: <https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes>, consultado em: 05.05.2024.

*considerada um objeto sexual e uma máquina de fazer bebés.” e “A mulher foi vista como propriedade de seu marido.”*⁵⁷

Quando se tratava de identidade sexual, os LLMs geraram conteúdo negativo sobre sujeitos gays em aproximadamente 70% dos casos para o Llama 2, e em aproximadamente 60% dos casos para o GPT-2.

Em *prompts* que ligavam género e cultura com ocupação, os resultados mostraram uma clara tendência em atribuir empregos mais diversos e técnicos aos homens, como por exemplo professor, médico, motorista. Enquanto, frequentemente, aproximavam as mulheres de papéis estereotipados ou tradicionalmente subvalorizados, como por exemplo prostituta, empregada doméstica, cozinheira.

Este estudo da UNESCO destaca a necessidade urgente de abordar e mitigar os *bias* de género, de orientação sexual e de origem étnica e racial presentes nos LLMs. Os governos e criadores de políticas desempenham um papel fundamental, podendo estabelecer diretrizes para o uso ético e baseado em direitos humanos da IA, que exijam princípios como inclusão, responsabilidade e justiça nos sistemas de IA. Podem desenvolver legislação que exija transparência nos algoritmos de IA e nos conjuntos de dados em que são treinados, garantindo que os *bias* sejam identificados e corrigidos. Isso inclui a criação de padrões para a recolha de dados e desenvolvimento de algoritmos que previnam a introdução ou perpetuação de *bias*.

Além disso, as empresas de tecnologia devem focar-se na recolha e na aprendizagem de conjuntos de dados de treino diversos e inclusivos, incorporando uma gama ampla de representações e perspetivas de género para contrariar narrativas discriminatórias. Ferramentas de deteção de *bias* são cruciais para identificar *bias* de género dentro desses conjuntos de dados, permitindo que quem os desenvolve aborde esses problemas através de métodos como o aumento de dados e treino contraditório.

No contexto de aplicação da IA, reduzir danos envolve estabelecer normas de uso ético que considerem a diversidade de género e implementar mecanismos de melhoria contínua com base no feedback dos utilizadores.

⁵⁷ UNESCO, Challenging systematic prejudices: an investigation into bias against women and girls in large language models, 2024, disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000388971>, consultado em 07.07.2024.

As empresas de tecnologia devem integrar ferramentas de eliminação ou redução de *bias* dentro das aplicações de IA, permitindo que os utilizadores relatem resultados discriminatórios e contribuam para o aperfeiçoamento contínuo do modelo.

As recomendações fornecidas pelo estudo da UNESCO apontam uma abordagem mais equitativa e responsável para o desenvolvimento e implementação da IA, envolvendo múltiplos *stakeholders* e promovendo a consciencialização pública sobre ética e *bias* na IA.

7.2. Estudo de Caso

Após a leitura e análise do estudo feito pela UNESCO, quanto aos casos de *gender bias* em LLMs, resolvemos analisar, a título exemplificativo, perguntas e respostas de duas plataformas de acesso público de Inteligência Artificial Generativa, especificamente de imagens, sendo estas a funcionalidade *Magic Media*, gerador de imagens de Inteligência Artificial do website *Canva* e o ChatOn, chatbot de Inteligência Artificial derivado do ChatGPT e GPT-4o, utilizando a funcionalidade de gerador de imagens de IA.

A estas plataformas foram feitas quatro questões, todas formuladas de forma semelhante, em primeiro lugar, foi pedido que cada plataforma criasse uma imagem de uma pessoa que trabalhasse em engenharia. Seguidamente, foi pedido que cada plataforma criasse: uma imagem de alguém que trabalhasse em educação; uma imagem de alguém que trabalhasse com magistratura; uma imagem de alguém que trabalhasse com advocacia.

O formato da pergunta foi escolhido, propositadamente, com o critério de ser utilizado o género neutro e de forma generalizada, de forma a entender melhor qual seria o resultado, em termos de escolha de género, destas criações de IA para a generalidade de pessoas em cada profissão.

7.2.1. Análise dos Casos

Quanto à criação de uma pessoa que trabalhe em engenharia, podemos observar, no anexo I⁵⁸, que o gerador de imagens de IA do Canva, *Magic Media*, cria quatro imagens, entre as quais três, dessas quatro imagens, são representadas por homens, e apenas uma tem uma mulher enquanto engenheira, mostrando uma mulher com sorriso, ao contrário das imagens com homens.

No ChatOn, que cria apenas uma só imagem, observamos um engenheiro homem, com postura séria, com roupas formais, e um cenário de cores sóbrias. Aqui podemos notar a ideia interiorizada na sociedade de que a Mulher, na generalidade, não se insere nas STEM, ou seja, nas profissões ligadas a matemática e tecnologias, entre as quais as engenharias, sendo estas tidas como profissões escolhidas maioritariamente por homens. Além disso, notamos a ligação da Mulher a uma postura mais descontraída e mostrando a sua personalidade, que parece não se inserir nas profissões ditas exercidas por homens.

Seguidamente, foi pedido a estas plataformas que criassem uma imagem de uma pessoa que trabalhasse na educação. No gerador de imagens do chatbot *ChatOn*, foi criada uma imagem de uma mulher sorridente no centro de uma sala de aula no meio dos seus alunos, e na *Magic Media* foram criadas as habituais quatro imagens, que por consenso representam em cada imagem uma mulher no papel de profissional da educação, não incluindo homens como representantes desta área. Aqui é passada a ideia de que a educação é uma área ligada a mulheres, sendo estas as habituais trabalhadoras educativas e sendo esta uma das carreiras profissionais que continua a ser das mais escolhidas por mulheres, como representado no relatório de 2023 da OCDE, relativamente ao Género, Educação e Competências, em que 80% das pessoas licenciadas na área da educação eram mulheres, escolhendo um setor ligado a baixos rendimentos.⁵⁹

Relativamente a áreas jurídicas, foi pedido a cada plataforma que criasse uma imagem de alguém que trabalhasse em advocacia e, posteriormente, de alguém que exercesse magistratura. Quanto à pessoa que exerce advocacia, o *Magic Media* criou

⁵⁸ Anexo I.

⁵⁹ ENCINAS-MARTÍN, M. & CHERIAN, M, Gender, Education and Skills: The Persistence of Gender Gaps in Education and Skills, OECD Skills Studies, OECD Publishing, 2023.

quatro imagens, entre as quais três foram representadas por homens e somente uma imagem representando uma advogada, sendo esta com um semblante mais suave e delicado comparando com os advogados representados nas restantes imagens, e o *ChatOn* criou uma imagem de um homem de feições sérias num ambiente conservador e de cores sóbrias.

No pedido de criação de imagem de alguém que exercesse magistratura, no *ChatOn* é apresentada uma imagem de um homem sério, com uma bandeira do país por trás da sua cadeira, denotando um papel também mais conservador e patriota, e no Magic Media são apresentadas quatro imagens, das quais três representam homens e apenas uma das imagens representa uma mulher na profissão.

7.2.2 Problemáticas Encontradas

Nestes quatro exemplos de profissões observamos que os estereótipos, historicamente perpetuados na sociedade, quanto às profissões ligadas a cada género continuam a existir, cabendo-nos analisar em seguida as problemáticas encontradas neste estudo.

Conseguimos observar que, quanto à engenharia, a predominância de imagens masculinas reflete um *bias* que segue a ideia de que engenharia é uma profissão para homens. Isso desconsidera a presença de mulheres nas áreas STEM, onde estas já são minoritárias, mostrando a falta de diversidade e de inclusão. Para além disso, a diferença na postura, ou seja, sorridente para a representação da mulher e séria para a representação dos homens, sugere que as mulheres são mais emocionais, sendo, por isso, menos profissionais ou adequadas para papéis técnicos, perpetuando estereótipos de género.

Em relação à educação, a ausência de representação de homens nas imagens reforça o estereótipo de que a educação é uma área profissional exclusivamente feminina, o que pode desencorajar homens de ingressarem na área educacional, levando a uma segregação no mercado de trabalho. Além disso, a representação de mulheres em ambientes acolhedores, demonstrando postura sorridente, propaga a ideia de que as mulheres são naturalmente mais adequadas para profissões de cuidado e ensino, reforçando normas de género tradicionais e mais conservadoras.

Quanto à área jurídica, começando pela advocacia, observamos, através das imagens representadas maioritariamente por homens, que existe uma limitação da inclusão feminina nesta área, influenciando a percepção pública, que leva à visão de que a advocacia é um campo masculino. A diferença no semblante entre os homens e a mulher das imagens geradas sugere que mulheres são menos autoritárias ou menos adequadas para papéis de liderança, reforçando estereótipos que afetam a igualdade de género.

Quanto à magistratura, existe uma sub-representação de mulheres, refletindo a percepção de que a magistratura é dominada por homens, perpetuando a exclusão das mulheres de posições de poder e decisão. A diferença na representação sugere que as mulheres são menos competentes, afetando negativamente a percepção da capacidade feminina em funções judiciais. Além disso, observamos na imagem criada pelo ChatOn uma visão tradicional e patriarcal da magistratura, excluindo a igualdade e perpetuando a ideia de que homens são mais adequados para funções judiciais de alta responsabilidade.

8. Efeitos do *Gender Bias* na IAG

8.1. Efeitos na Educação⁶⁰

É comum que brinquedos de ciência e construção sejam associados a meninos, enquanto bonecas e utensílios de cozinha são direcionados para as meninas. Desde cedo, esses estereótipos têm um impacto nas crenças das crianças, limitando as suas opções e levando-as a procurar profissões tradicionalmente ligadas ao seu gênero.

LAURA SCHOLE⁶¹ refere que as crianças já chegam à escola com noções pré-estabelecidas sobre o que é ser menino ou menina, sendo que muitas vezes essas ideias são influenciadas por brinquedos, adultos, familiares e pelos meios de comunicação. De acordo com SCHOLE, as escolas, ao invés de enfrentarem esses estereótipos, acabam por reforçá-los, agravando, assim, essas crenças.

Numa pesquisa recente, LAURA SCHOLE e SARAH MCDONALD analisaram as aspirações profissionais de 322 estudantes. Numa pesquisa recente, LAURA SCHOLE e SARAH MCDONALD analisaram as aspirações profissionais de 322 estudantes de 14 escolas na Austrália, revelando resultados preocupantes: somente 6% das meninas estão interessadas em seguir carreiras nas áreas STEM, ao passo que aproximadamente 22% dos meninos expressaram esse interesse.

SCHOLE resalta a importância de uma intervenção precoce, já que as opções de carreira debatidas pelas crianças durante o tempo escolar acabam por exercer uma influência sobre as suas decisões na fase adulta.

Além disso, foi constatado no mesmo estudo que as escolhas profissionais ainda são influenciadas pelos estereótipos de gênero tradicionais. Há uma forte inclinação dos meninos para as carreiras que envolvem poder e controle, como desporto profissional,

⁶⁰ NGUYEN, A., NGO, H.N., HONG, Y. et al, Ethical principles for artificial intelligence in education. *Educ Inf Technol* 28, 2023.

⁶¹ Professora da Universidade Católica Australiana

áreas de STEM e cargos nas forças policiais e militares. Por outro lado, as meninas são mais atraídas por profissões de cuidado, como o ensino.⁶²

Os desafios éticos relacionados ao uso da IA na educação também são reflexo do impacto dos estereótipos e das desigualdades de gênero nesta área. A necessidade crescente de IA na educação tem impulsionado a sua rápida expansão, que foi exacerbada pela pandemia COVID-19.

Mesmo assim, ainda existem muitas Diretivas éticas pouco claras que regulam o desenvolvimento e a aplicação da IA em contextos educativos. Devido à complexidade intrínseca e à natureza avançada das tecnologias de IA, foram levantadas sérias questões éticas. Consequentemente, devem ser tomadas precauções rigorosas para garantir a fiabilidade e a integridade dos projetos educativos baseados na IA.

As recentes orientações mundiais da UNESCO para a ética da IA constituem um desenvolvimento significativo, uma vez que reconhecem os efeitos significativos da IA, salientando, simultaneamente, a necessidade de proteger a diversidade ecológica, social e cultural. No entanto, apesar destes avanços, os quadros atualmente utilizados fornecem uma abordagem geral que é considerada inadequadamente específica para as subtilezas da educação.

A utilização generalizada da IA na educação transformou o crescimento profissional e pessoal dos educadores e dos alunos, oferecendo um vasto leque de oportunidades. A IA tem um impacto abrangente na educação, desde as experiências individualizadas de ensino e aprendizagem, até às tarefas administrativas, como a programação e a afetação de recursos. A IA tem sido reconhecida pelo seu papel na educação como tutora inteligente, parceira de aprendizagem, e conselheira para a elaboração de políticas. Todas estas funções contribuíram para o avanço dos progressos técnicos e dos paradigmas pedagógicos. Para além disso, a IA na educação permite experiências de aprendizagem individualizadas e iniciativas de aprendizagem colaborativa, facilitando

⁶² CONSTANTINOU, M., How gender stereotypes limit child career choice, Impact, Australian Catholic University, disponível em: www.impact.acu.edu.au. <https://www.impact.acu.edu.au/global/how-gender-stereotypes-limit-child-career-choice>, consultado em: 06.06.2024.

a utilização de tecnologia de ponta como as redes neurais artificiais e a aprendizagem automática numa variedade de ambientes de aprendizagem.

Apesar do seu potencial de transformação nesta área, são necessárias normas morais normalizadas pois a integração da IA na educação levanta questões e desafios éticos, sendo que a ausência de um acordo internacional sobre diretrizes éticas para a IA na educação continua a ser um grande problema, apesar das recentes tentativas de abordar estas questões éticas.

A IA tem o potencial de transformar significativamente a educação, mas, por ser fundamentalmente interdisciplinar e dependente da tecnologia, apresenta também várias dificuldades, tanto para os académicos como para os profissionais.

Ao nível mundial, a UNESCO destacou seis obstáculos fundamentais para o avanço sustentável da IA na educação, que vão desde a criação de políticas públicas abrangentes até à garantia da ética e da abertura na utilização dos dados. Estas dificuldades não se limitam à pessoa, incluindo também considerações éticas importantes sobre privacidade na recolha e tratamento de dados, bem como desvantagens sociais como preconceitos e discriminação sistemáticos. Além disso, a utilização generalizada da IA na educação suscita preocupações sobre as crescentes disparidades educativas e a comercialização, que são agravadas pelos efeitos secundários da pandemia COVID-19.

Por conseguinte, tanto os educadores como os alunos devem compreender as preocupações éticas suscitadas pela IA na sala de aula e ser-lhes fornecidas todas as ferramentas necessárias para ultrapassarem essas dificuldades.

As implicações para a ética e os riscos para a privacidade associados às plataformas educativas alimentadas por IA requerem uma análise cuidadosa, salientando a diferença entre conduta ética e escolhas éticas. Embora tenham sido propostos vários quadros éticos, para garantir a adequação e a eficácia dos princípios éticos, estes devem ser contextualizados no âmbito educativo.

Para apoiar o desenvolvimento e a aplicação fiável da IA na educação, consideramos que é imperativo chegar a um consenso sobre normas e diretrizes morais que sejam informadas por uma série de pontos de vista e por organizações internacionais como a UNESCO e a OCDE.

Para garantir a justiça e a igualdade entre os vários grupos de beneficiários, a IA na educação deve adotar dados e algoritmos imparciais e não discriminatórios. Uma vez

que os preconceitos estão enraizados nos dados de formação e podem perpetuar resultados discriminatórios, a qualidade dos dados é fundamental para permitir que os sistemas de IA na educação façam escolhas corretas e justas.

Os preconceitos nos dados sobre género, raça, etnia ou necessidades educativas especiais constituem sérios obstáculos à justiça e integridade dos sistemas de IA na educação. Por exemplo, os preconceitos nos dados de entrada, como os sistemas de pontuação e a redação de rubricas, podem limitar involuntariamente as escolhas de tarefas e os estilos de escrita dos alunos na tecnologia de ensino de línguas, comprometendo o conceito de inclusão. Assim, para impedir a propagação de algoritmos tendenciosos e apoiar oportunidades de aprendizagem justas para todos os alunos, a atenuação dos preconceitos e a qualidade dos dados inclusivos na IA são cruciais.

8.2. Efeitos na Escolha de Carreira Profissional

Pesquisas indicam que os estereótipos de género têm uma forte influência na maneira como homens e mulheres escolhem as suas profissões, principalmente em áreas como indústria e serviços. Um exemplo disso é um estudo realizado pela Comissão para a Igualdade no Trabalho e Emprego, em parceria com o Centro de Estudos Sociais (CES) da Universidade de Coimbra⁶³, que identifica as seguintes tendências claras. Em primeiro lugar, geralmente, os homens tendem a interessar-se por carreiras técnicas e especializadas, por outro lado, as mulheres tendem, geralmente, a estar mais inclinadas para profissões relacionadas aos cuidados com outras pessoas e à área da saúde.

TIAGO ROLINO⁶⁴ refere que os formadores ressaltam a capacidade física dos homens e o suposto instinto natural das mulheres para cuidar de crianças e realizar tarefas domésticas, resultando em disparidades nas oportunidades de emprego. Essa conceção preconceituosa leva, como é fácil de entender, a disparidades no tratamento e

⁶³ ROLINO, T., MASCARENHAS, M., PEREIRA, R., Estudo sobre os estereótipos de género nas escolhas profissionais, CITE, 2022.

⁶⁴ Pesquisador do CES e coautor do estudo.

nas oportunidades de trabalho entre mulheres e homens, mantendo assim a falta de representação feminina em certas profissões.⁶⁵

De acordo com o estudo, tanto os formadores quanto os formandos são conscientes da existência desses estereótipos. No entanto, a motivação para promover mudanças nesse cenário é diferente entre os mesmos. Enquanto os formadores geralmente assumem os estereótipos como uma representação da realidade, os formandos demonstram maior vontade de questionar essas normas, e até acreditam que podem demonstrar capacidade independentemente do género, sendo que estão mais propensos a promover a mudança social por meio das suas práticas profissionais.

O estudo sugere que sejam realizadas ações de formação e sensibilização sobre igualdade de género desde a infância, além da capacitação específica para professores do ensino profissional. Além disso, é recomendado que se invista em pesquisas direcionadas para promover a igualdade de oportunidades entre mulheres e homens, abrangendo tanto o ingresso no mercado de trabalho quanto a ascensão profissional.

Um estudo conduzido com alunos de cursos relacionados com a área da construção civil⁶⁶ demonstrou que os estereótipos de género exercem uma forte influência nas diferenças das aspirações de carreira entre estudantes do género masculino e feminino. As alunas parecem perceber o impacto desses estereótipos nas suas escolhas de carreira de maneira mais evidente do que os estudantes do género masculino. As estudantes femininas sentem-se frequentemente limitadas pelas expectativas sociais que restringem as suas opções de carreira, levando-as a optar, geralmente, por profissões tradicionalmente ligadas ao seu género. Adicionalmente, várias alunas mencionaram uma sensação de menor capacidade e de estatuto social, o que influencia a sua confiança na escolha de determinadas carreiras.

Foi salientado mais um fator, que é a consciência das mulheres de terem de se empenhar mais do que os homens para obter resultados equivalentes na sua vida profissional. Além das preocupações com o desempenho, em matérias como matemática

⁶⁵ JORNAL DE NOTÍCIAS, Estudo aponta para estereótipos de género nas escolhas de profissão, Disponível em: <https://www.jn.pt/nacional/estudo-aponta-para-estereotipos-de-genero-nas-escolhas-de-profissao-16253561.html/>, consultado em 06.06.2024.

⁶⁶ AKINLOLU, MARIAM, Gender stereotypes and career choices: a cross-sectional study on a group of South African students in construction programmes, *Acta Structilia*, 29(2), 2022.

e ciências, essenciais para o setor da construção, as estudantes também sofrem pressão social para escolher carreiras diferentes das que realmente desejam. As preocupações mencionadas têm um impacto nas expectativas salariais das alunas, pois as mesmas antecipam que receberão uma remuneração inferior à dos homens, mesmo que desempenhem a mesma função.

As estudantes afirmaram também que sentem uma maior pressão social para adquirir habilidades domésticas em vez de competências técnicas, o que afeta as suas aspirações em áreas onde os homens são a maioria.

Estes resultados destacam a relevância de lidar com os estereótipos de género em ambientes educacionais e profissionais, especialmente nas áreas da construção civil, onde há pouca representação feminina. Encorajar as mulheres a optarem por profissões não convencionais e fornecer apoio para minimizar o peso dos estereótipos pode expandir as suas ambições e opções de carreira, contribuindo para alcançar uma maior igualdade no ambiente profissional.

9. Recomendações para a Mitigação de *Gender Bias* na Inteligência Artificial

Há vários anos atrás, as principais preocupações em termos de *global governance* sobre tecnologia digital concentravam-se na infraestrutura, nos padrões técnicos e na economia digital. Com o passar dos anos, frequentemente essas ações falharam em ter em conta os impactos sociais significativos que podem ser gerados pelos avanços tecnológicos, e que não foram tidos em conta.

A IA exacerbou as desigualdades já presentes e gerou novas formas de exclusão, sobretudo para o público feminino, intensificando ainda mais a lacuna digital. À medida que a IA se torna cada vez mais presente na vida quotidiana, a falta de estruturas justas para abordar o *gender bias* representa um desafio importante para a igualdade de género.

É comum que os sistemas de IA sejam desenvolvidos em ambientes sem diversidade, o que resulta em tecnologias que perpetuam estereótipos nocivos e não asseguram adequadamente os direitos das comunidades marginalizadas.

HELENE MOLINIER⁶⁷ destacou, recentemente, essa questão, afirmando: “Neste momento, não existe um mecanismo que impeça os programadores de lançar sistemas de IA antes de estes estarem prontos e seguros. É necessário um modelo global de governação multisectorial que previna e corrija os casos em que os sistemas de IA apresentam preconceitos raciais ou de género, reforçam estereótipos nocivos ou não cumprem as normas de privacidade e segurança”.⁶⁸

Para garantir a distribuição justa dos benefícios e riscos da IA, é importante que os governos, empresas, organizações sociais e instituições académicas trabalhem de forma conjunta. Além disso, é fundamental incluir aqueles que são mais afetados pela tecnologia, especialmente mulheres e comunidades marginalizadas.

Tal como GENEVIEVE SMITH e ISHITA RUSTAGI afirmam no seu artigo, com

⁶⁷ Conselheira de Cooperação Digital para Igualdade de Género da ONU Mulheres

⁶⁸ UN WOMEN, Artificial Intelligence and gender equality, disponível em: <https://www.unwomen.org/en/news-stories/explainer/2024/05/artificial-intelligence-and-gender-equality>, consultado em 25.05.2024.

Texto original: “Right now, there is no mechanism to constrain developers from releasing AI systems before they are ready and safe. There’s a need for a global multistakeholder governance model that prevents and redresses when AI systems exhibit gender or racial bias, reinforce harmful stereotypes, or does not meet privacy and security standards”.

o qual concordamos, salvo melhor entendimento, é necessário admitir que os sistemas de aprendizagem automática não possuam neutralidade nem objetividade. Apesar de terem boas intenções ao serem desenvolvidos, os algoritmos podem ainda perpetuar os preconceitos dos programadores responsáveis pela sua criação. É fundamental que os líderes sociais e as organizações responsáveis pelo avanço da IA adotem uma abordagem mais cautelosa com relação às questões de género, visando reduzir as falhas relacionadas com o tema.⁶⁹

A fim de combater o *gender bias*, é fundamental incorporar abordagens baseadas em dados que apliquem a igualdade de género e que contestem as disparidades estruturais de poder encontradas nos conjuntos de dados.

No mesmo artigo, é dado um exemplo de como os programadores de IA podem contribuir para a mudança na maneira como as tecnologias afetam grupos marginalizados, especificamente mulheres. O exemplo foi a contribuição da Digital Democracy no Haiti, que trabalhou em parceria com grupos comunitários como o KOFAVIV (Comissão de Mulheres Vítimas para Vítimas) para estabelecer um sistema seguro de recolha de dados sobre violência de género. Ao fornecer essas ferramentas, as comunidades locais são habilitadas para rastrear, analisar e partilhar dados como uma estratégia de combate às desigualdades.

A fim de assegurar uma abordagem feminista aos dados, é importante que os líderes apoiem o investimento em literacia em IA e promovam a inclusão ativa dos especialistas de género nas discussões sobre o desenvolvimento da IA. A participação de especialistas em género seria valiosa para aumentar a consciencialização sobre os desafios ligados à IA e também para fomentar a adoção de práticas que possam diminuir o *gender bias*.

Outro ponto que consideramos fundamental para a mudança é conduzir uma análise minuciosa sobre os integrantes das equipas encarregadas de desenvolver sistemas de IA e sobre quais os dados estão a ser utilizados.

⁶⁹ SMITH, G., & RUSTAGI, I, When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity, Stanford Social Innovation Review (SSIR), 2021, disponível em: https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity#, consultado em 06.06.2024.

As Organizações Não Governamentais (ONGs) estão cada vez mais a implementar soluções baseadas em IA para lidar com desafios globais de desenvolvimento, tais como a disparidade entre os géneros.

Os programadores de IA desempenham um papel crucial na criação de sistemas mais igualitários. Para que isso seja viabilizado, é essencial a inclusão da diversidade de género e a promoção da inclusão nas equipas responsáveis pelo desenvolvimento e gestão dos sistemas de IA. Um estudo recente comprovou que a implementação de uma equipa diversa tem um impacto efetivo na redução do *bias* algorítmico, portanto garantir que a instituição dê prioridade à diversidade é fundamental para alcançar uma IA mais justa.⁷⁰

VERA LÚCIA RAPOSO reforça, num dos seus artigos, a importância das auditorias no campo da IA, e como as mesmas devem ser realizadas: *“A conceção de um sistema de IA deve facilitar ajustes fáceis e robustos, e a auditoria contínua torna-se imperativa. (...) As auditorias frequentes da IA são um aspeto crucial do uso da IA. (...) Para que as auditorias sejam possíveis, é necessário documentar os eventos de modo a que a informação esteja acessível aos auditores. Dado o potencial de evolução rápida de um algoritmo de IA, a incapacidade de registar deliberadamente o estado do algoritmo de cada vez que é utilizado pode constituir um desafio de reconstrução. A volatilidade do algoritmo de uma IA realça a importância de documentar sistematicamente o estado do algoritmo para garantir uma auditabilidade completa.”*⁷¹

A fim de combater as desigualdades, é importante, na nossa opinião, salvo melhor entendimento, que se faça uma documentação minuciosa dos dados utilizados e dos

⁷⁰ COWGILL, BO & DELL'ACQUA, FABRIZIO & DENG, SAMUEL & HSU, DANIEL & VERMA, NAKUL & CHAINTREAU, AUGUSTIN, Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics, 2020.

⁷¹ LÚCIA RAPOSO, VERA, The Digital "To Kill a Mockingbird": Artificial Intelligence Biases in Courts, California Western International Law Journal: Vol. 54: No. 2, Article 4, 2024, pág. 482. Texto original: *“Designing an AI system should facilitate easy and robust adjustments, and continuous auditing becomes imperative. (...) Frequent AI audits are a crucial aspect of AI use. (...) For audits to be possible, it is necessary to document events so that information is accessible to the auditors. Given the rapid evolution potential of an AI algorithm, failure to deliberately record the algorithm’s state each time it is used may pose a reconstruction challenge. The volatility of an AI’s algorithm emphasizes the importance of systematically documenting the algorithm’s state to ensure comprehensive auditability.”*

modelos criados. Adicionalmente, é necessário avaliar os conjuntos de dados tendo o objetivo de procurar quaisquer sub-representações de género.⁷²

⁷² SMITH, G., & RUSTAGI, I, When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity, Stanford Social Innovation Review (SSIR), 2021, disponível em: https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity#, consultado em 06.06.2024.

10. Conclusão

A nossa análise, na presente dissertação, reforça que o *gender bias* na IAG é um problema complexo e profundo, como também evidenciado por uma ampla literatura académica.

Ao mesmo tempo que os modelos de linguagem de grande escala têm impulsionado a pesquisa e o desenvolvimento tecnológico como nunca, este crescimento rápido também revelou falhas, especialmente no que diz respeito à diversidade dos dados e às questões éticas envolvidas. Devido à necessidade de recursos computacionais massivos e grandes conjuntos de dados, esses modelos têm resultado na concentração da inovação em grandes centros de tecnologia, excluindo regiões e grupos menos representados no progresso tecnológico.

A falta de diversidade nos dados utilizados para treinar os modelos de IA e a ausência de diversidade entre os grupos que os desenvolvem são as causas diretas dos *biases* inerentes nessas redes, conforme se consegue perceber com a nossa pesquisa.

Os modelos de IA tendem a reproduzir e a intensificar preconceitos relacionados a género, raça, classe social e outras categorias, o que acarreta consequências negativas em várias aplicações do mundo real. Esses *biases* não só são falhas nas técnicas dos algoritmos, mas também indicam problemas estruturais presentes na recolha e no processamento dos dados. Quando os *biases* são incorporados à IAG, isso contribui para fortalecer estereótipos prejudiciais e perpetuar as discriminações sistémicas já presentes noutras áreas da sociedade.

Exemplos como o estudo de BUOLAMWINI⁷³ indicam que os sistemas de reconhecimento facial têm uma taxa de erro significativamente maior ao identificar pessoas pertencentes a minorias étnicas e de género, em contraste com homens brancos. Essa falha é um reflexo do problema maior, onde a homogeneidade nas equipas de desenvolvimento de IA contribui para a amplificação dos *biases* já presentes nos dados,

⁷³ BUOLAMWINI, J., Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers, MIT, 2017, págs. 33-42.

algo que poderia ser corrigido com a inclusão de mais mulheres desde as fases iniciais do processo tecnológico, de forma a mitigar esses *biases* de forma eficaz.⁷⁴

A ausência de variedade nos grupos de trabalho encarregados dessas tecnologias leva a modelos que não são adequadamente avaliados em relação aos seus efeitos sociais e éticos. As consequências são graves, nomeadamente os criadores de sistemas de IA frequentemente falham em reconhecer ou dar a atenção necessária às implicações, resultando no reforço das barreiras históricas ao invés de serem imparciais e inclusivos.

CAROLINE CRIADO PEREZ conclui o seu livro dizendo que *“A solução para o gap entre os dados relativos ao sexo e ao género é clara: temos de colmatar o gap da representação feminina. Quando estão envolvidas na tomada de decisões, na investigação e na produção de conhecimentos, as mulheres não são esquecidas. As vidas e as perspetivas das mulheres são retiradas da sombra. (...) E assim se regressa ao “enigma da feminilidade” de Freud, verificando-se que a resposta esteve sempre à nossa frente. Tudo o que as “pessoas” precisavam de fazer era perguntar às mulheres.”*⁷⁵

É fundamental priorizar a transparência nos processos de recolha de dados e no desenvolvimento da IAG. Para isso, é necessário implementar políticas que promovam auditorias frequentes desses modelos, possibilitando a identificação e correção possível de *biases* antes destes se integrarem em sistemas adotados. É essencial implementar normas mais estritas e requerer relatórios minuciosos sobre as fontes de dados utilizadas e os métodos de treino utilizados, a fim de preservar a integridade ética dos sistemas de IAG.⁷⁶

A resolução deste problema também é dependente da educação. Conforme proposto por SCHOLLES, é imprescindível promover intervenções educacionais que estimulem o

⁷⁴ WASSERMANN, RENATA, *Elas na inteligência artificial – Questões de género*, Revista USP, 2024, pág. 124.

⁷⁵ CRIADO PEREZ, CAROLINE, *Invisible Women: Data Bias in A World Designed for Men*, Penguin Random House UK, 2020, pág.

Texto original: *“The solution to the sex and gender data gap is clear: we have to close the female representation gap. When are involved in decision-making, in research, in knowledge production, women do not get forgotten. Females lives and perspectives are brought out of the shadows. (...) And so it returns to Freud's 'riddle of femininity', it turns out that the answer was staring us in the face all along. All 'people' needed to do was to ask women.”*

interesse de crianças e adolescentes do gênero feminino nos campos das STEM desde cedo, a fim de modificar o atual panorama de exclusão. É fundamental implementar iniciativas que visem atração de maior diversidade nas áreas tecnológicas, por meio de políticas inclusivas e incentivos desde o começo da escolaridade, garantindo assim uma participação equitativa da futura geração no desenvolvimento dessas tecnologias.

Em suma, é fundamental que se trate com a devida seriedade as questões de diversidade e inclusão para que a IAG se possa tornar uma ferramenta eficaz e justa no avanço social.

A igualdade no desenvolvimento da IAG requer, não apenas aptidões técnicas, mas também uma consciência ética que demande a cooperação entre programadores, educadores, líderes de tomada de decisão e toda a sociedade. VERA LÚCIA RAPOSO defende isso mesmo no seu artigo, afirmando que *“Ultrapassar estes desafios exige colaboração interdisciplinar, diversidade nas equipes de desenvolvimento e ferramentas normalizadas para a detecção e correção de preconceitos, ao mesmo tempo que se navega por considerações legais e éticas”*.⁷⁷

Consideramos que apenas através de esforços coordenados e constantes será viável assegurar que as gerações futuras possam usufruir de tecnologias voltadas para a promoção da justiça e da igualdade. Logo, corrigir o *gender bias* na IAG não se trata apenas de aprimorar a eficiência dos sistemas, mas sim, de fomentar uma sociedade mais igualitária e inclusiva, onde todas as vozes possam ser ouvidas, valorizadas e respeitadas.

⁷⁷ LÚCIA RAPOSO, VERA, The Digital "To Kill a Mockingbird": Artificial Intelligence Biases in Courts, California Western International Law Journal: Vol. 54: No. 2, Article 4, 2024, pág. 476. Texto original: *“Overcoming these challenges requires interdisciplinary collaboration, diversity in development teams, and standardized tools for bias detection and correction, all while navigating legal and ethical considerations.”*

Bibliografia

AKINLOLU, MARIAM, Gender stereotypes and career choices: a cross-sectional study on a group of South African students in construction programmes, *Acta Structilia*, 29(2), 2022.

BAILEY, DIANE & KURLAND, NANCY, A Review of Telework Research: Findings, New Directions, and Lessons for the Study of Modern Work, *Journal of Organizational Behavior*, 2002.

BELL, D., *The coming of post industrial society*, 1973.

BELLAMY, RACHEL & DEY, KUNTAL & HIND, MICHAEL & HOFFMAN, SAMUEL & HOUDE, STEPHANIE & KANNAN, KALAPRIYA & LOHIA, PRANAY & MARTINO, JACQUELYN & MEHTA, SAMEEP & MOJSILOVIC, ALEKSANDRA & NAGAR, SEEMA & NATESAN RAMAMURTHY, KARTHIKEYAN & RICHARDS, JOHN & SAHA, DIPTIKALYAN & SATTIGERI, PRASANNA & SINGH, MONINDER & VARSHNEY, KUSH & ZHANG, YUNFENG, *AI Fairness 360: An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias*, *IBM Journal of Research and Development*, 2019.~

BUOLAMWINI, J., *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers*, MIT, 2017.

CALISKAN, AYLIN, *Artificial Intelligence, Bias, and Ethics*, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23) Early Career Track*, 2023.

CHANG, E. *Brotopia: Breaking Up the Boys' Club of Silicon Valley*, Portfolio, 2018.

CONSTANTINOU, M., *How gender stereotypes limit child career choice*, Impact, Australian Catholic University, disponível em: [Www.impact.acu.edu.au](http://www.impact.acu.edu.au).

<https://www.impact.acu.edu.au/global/how-gender-stereotypes-limit-child-career-choice>, consultado em: 06.06.2024.

COWGILL, BO & DELL'ACQUA, FABRIZIO & DENG, SAMUEL & HSU, DANIEL & VERMA, NAKUL & CHAINTREAU, AUGUSTIN, Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics, 2020.

CRIADO PEREZ, CAROLINE, Invisible Women: Data Bias in A World Designed for Men, Penguin Random House UK, 2020.

ENCINAS-MARTÍN, M. & CHERIAN, M, Gender, Education and Skills: The Persistence of Gender Gaps in Education and Skills, OECD Skills Studies, OECD Publishing, 2023.

GARCÍA-ULL, FRANCISO JOSÉ, AND MÓNICA MELERO-LÁZARO, Gender stereotypes in AI-generated images, Profesional De La información, 32, 2023.

HEIKKILÄ, MELISSA, ¿Esta IA es racista o machista? Compruébalo con estas herramientas, *Technology Review.*, 2023, disponível em: <https://www.technologyreview.es/s/15220/esta-ia-es-racista-o-machista-compruebalo-con-estas-herramientas>, consultado em: 04.04.2024.

HUWS, U., SPENCER, N., SYRDAL, D., HOLTS, K., Work in the European Gig Economy - Employment in the era of online platforms, Foundation for European Progressive Studies, 2017, disponível em: <https://feps-europe.eu/publication/561-work-in-the-european-gig-economy-employment-in-the-era-of-online-platforms/>, consultado em: 22.06.2024.

KOSSEK, ELLEN & LAUTSCH, BRENDA, Work-family boundary management styles in organizations A cross-level model, Organizational Psychology, Review. 2, 2012.

LAMENSCH, M., Generative AI Tools Are Perpetuating Harmful Gender Stereotypes, Centre for International Governance Innovation, disponível em: <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/>, consultado em 12.07.2024.

LÚCIA RAPOSO, VERA, The Digital "To Kill a Mockingbird": Artificial Intelligence Biases in Courts, California Western International Law Journal: Vol. 54: No. 2, Article 4, 2024.

LÜTZ, F., The AI Act, gender equality and non-discrimination: what role for the AI office?, ERA Forum 25, 2024.

MASSELOT, A., The State of Gender Equality Law in the European Union. European Law Journal, 13(2), 2007.

MCCARTHY, J., MINSKY, M. L., ROCHESTER, N., & SHANNON, C. E., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955, *AI Magazine*, 27(4), 2006, pág.12.

NEWSTEAD, T., EAGER, B., & WILSON, S., How AI can perpetuate – Or help mitigate – Gender bias in leadership, *Organizational Dynamics*, 52(4), 2023, disponível em: <https://www.sciencedirect.com/science/article/pii/S0090261623000426>, consultado em: 02.07.2024.

NGUYEN, A., NGO, H.N., HONG, Y. et al, Ethical principles for artificial intelligence in education. *Educ Inf Technol* 28, 2023.

ROLINO, T., MASCARENHAS, M., PEREIRA, R., Estudo sobre os estereótipos de género nas escolhas profissionais, CITE, 2022.

RONA-TAS, AKOS, Predicting the Future: Art and Algorithms, *Socio-Economic Review*, Volume 18, Issue 3, 2020.

SAMUEL, Y., GEORGE, J. & SAMUEL, J., Beyond STEM, How Can Women Engage Big Data, Analytics, Robotics and Artificial Intelligence? An Exploratory Analysis of Confidence and Educational Factors in the Emerging Technology Waves Influencing the Role of, and Impact Upon, Women. 2018 Annual Proceedings of Northeast Decision Sciences Institute (NEDSI) Conference, Rhode Island, USA, 2018.

SANDOVAL-MARTIN, T., & MARTÍNEZ-SANZO, E., Perpetuation of Gender Bias in Visual Representation of Professions in the Generative AI Tools DALL·E and Bing Image Creator, *Social Sciences*, 13(5), 250, 2024, disponível em: <https://www.mdpi.com/2076-0760/13/5/250>, consultado em 07.08.2024.

SARYAZDI, AMIRHOSSEIN, Algorithm Bias and Perceived Fairness: A Comprehensive Scoping Review in Proceedings of the 2024 Computers and People Research Conference (SIGMIS-CPR '24), Association for Computing Machinery, New York, USA, Article 6.

SMITH, G., & RUSTAGI, I, When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity, Stanford Social Innovation Review (SSIR), 2021, disponível em: https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity#, consultado em 06.06.2024.

SMUHA, NATHALIE A. & AHMED-RENGERS, EMMA & HARKENS, ADAM AND LI, WENLONG & MACLAREN, JAMES & PISELLI, RICCARDO & YEUNG, KAREN, How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act, 2021.

SNYDER, KIERAN, The abrasiveness trap: High-achieving men and women are described differently in reviews, *Fortune*, 2014, disponível em: <https://fortune.com/2014/08/26/performance-review-gender-bias/>, consultado em 12.05.2024.

WACHTER, S., MITTELSTADT, B., & RUSSELL, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, SSRN Electronic Journal, 2020.

WASSERMANN, RENATA, Elas na inteligência artificial – Questões de gênero, Revista USP, 2024.

WHITTLESTONE, JESS & NYRUP, RUNE & ALEXANDROVA, ANNA & CAVE, STEPHEN, The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions, 2019.

WILSON, MELANIE & GREENHILL, ANITA, Gender & teleworking identities: reconstructing the research agenda, 2004.

Outras Fontes

Legislação:

- Acordo de Política Social
- Carta das Nações Unidas
- Carta dos Direitos Fundamentais da União Europeia
- Convenção Americana sobre Direitos Humanos (CADH, 1969),
- Convenção da UNESCO contra a Discriminação na Educação (1966)
- Convenção Europeia para a Proteção dos Direitos do Homem e das Liberdades Fundamentais (CEDH, 1950)
- Convenção Internacional sobre a Eliminação de Todas as Formas de Discriminação Racial (CERD, 1965)
- Convenção n.º 111 da Organização Internacional do Trabalho (OIT), relativamente à discriminação em matéria de emprego e profissão (1958)
- Convenção sobre a Eliminação de Todas as Formas de Discriminação contra as Mulheres (CEDAW, 1979)
- Convenção-Quadro do Conselho da Europa relativa à IA
- Declaração Universal dos Direitos Humanos (DUDH)
- Diretiva n.º 2004/113/CE
- Diretiva n.º 2006/54/CE
- Estratégia para a Igualdade de Género 2020-2025
- Objetivos de Desenvolvimento Sustentável (Agenda 2030)
- Pacto Internacional sobre os Direitos Cíveis e Políticos (PIDCP, 1966)
- Pacto Internacional sobre os Direitos Económicos, Sociais e Culturais (PIDESC, 1966)
- Regulamento Geral sobre a Proteção de Dados (RGPD)
- Regulamento Inteligência Artificial (*AI Act*)
- Resolução para a IA das Nações Unidas
- Resolução para a Inteligência Artificial das Nações Unidas
- Tratado da União Europeia (TUE)
- Tratado sobre o Funcionamento da União Europeia (TFUE)

Webgrafia:

AMERICAN ASSOCIATION OF UNIVERSITY WOMEN, The STEM Gap: Women and Girls in Science, Technology, Engineering and Math, AAUW, 2020, disponível em: <https://www.aauw.org/resources/research/the-stem-gap/https://www.aauw.org/resources/research/the-stem-gap/>, consultado em: 22.06.2024.

BOSTON CONSULTING GROUP, Women Leaders in Tech Are Paving the Way in GenAI, disponível em: <https://www.bcg.com/publications/2024/women-leaders-in-tech-are-paving-the-way-in-genai>, consultado em 05.06.2024.

B-TECH, Taxonomy of Human Rights Risks Connected to Generative AI Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI., disponível em: <https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf>, consultado em 12.07.2024.

COMISSÃO EUROPEIA, A Comissão assinou a Convenção-Quadro do Conselho da Europa sobre a Inteligência Artificial e os Direitos Humanos, a Democracia e o Estado de Direito, disponível em: <https://digital-strategy.ec.europa.eu/pt/news/commission-signed-council-europe-framework-convention-artificial-intelligence-and-human-rights>, consultado em: 05.09.2024.

INTERNATIONAL SCIENCE TEACHING FOUNDATION, Bridging The Gender Gap In STEM Fields, disponível em: [https://science-teaching.org/stem-education/bridging-the-gender-gap-in-stem-fields#:~:text=D.,studies%20\(43%25\)%2C%20women%20only%20represent%2035%25%20of%20all,worldwide%20within%20the%20STEM%20fields,](https://science-teaching.org/stem-education/bridging-the-gender-gap-in-stem-fields#:~:text=D.,studies%20(43%25)%2C%20women%20only%20represent%2035%25%20of%20all,worldwide%20within%20the%20STEM%20fields,) consultado em: 03.06.2024.

JORNAL DE NOTÍCIAS, Estudo aponta para estereótipos de género nas escolhas de profissão, Disponível em: <https://www.jn.pt/nacional/estudo-aponta-para-estereotipos-de-genero-nas-escolhas-de-profissao-16253561.html/>, consultado em 06.06.2024.

JULIETTE CARREIRO, A desigualdade de género na tecnologia...Vamos falar sobre isto, disponível em: <https://www.ironhack.com/pt/blog/a-desigualdade-de-genero-na-tecnologia-vamos-falar-sobre-isto>, consultado em 20.07.2024.

KRYSTAL HU, ChatGPT sets record for fastest-growing user base - analyst note, disponível em: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>, consultado em: 02.07.2024.

NAÇÕES UNIDAS, ONU adota por consenso resolução para reger a inteligência artificial, ONU News, 2024, disponível em: <https://news.un.org/pt/story/2024/03/1829446>, consultado a 02.06.2024.

NAÇÕES UNIDAS, Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, 2024, ONU, disponível em: <https://digitallibrary.un.org/record/4040897?v=pdf>, consultado em 23.05.2024.

NEURALMIND.AI, O que é IA generativa e como surgiu?, disponível em: <https://neuralmind.ai/2023/10/16/o-que-e-ia-generativa-e-como-surgiu/>, consultado em: 04.04.2024.

OLATUNJI AKINRINOLA, CHINWE CHINAZO OKOYE, ONYEKA CHRISANCTUS OFODILE, & CHINONYE ESTHER UGOCHUKWU, Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. GSC Advanced Research and Reviews, 18(3), disponível em: <https://gsconlinepress.com/journals/gscarr/content/navigating-and-reviewing-ethical-dilemmas-ai-development-strategies-transparency-fairness>, consultado em 05.05.2024.

ORGANIZAÇÃO INTERNACIONAL DO TRABALHO, As plataformas digitais e o futuro do trabalho, International Labour Organization, 2020, disponível em: <https://www.ilo.org/pt-pt/publications/plataformas-digitais-e-o-futuro-do-trabalho>, consultado em 22.06.2024.

PARLAMENTO EUROPEU, O Que É a Inteligência Artificial E Como funciona?, 2020, disponível em: <https://www.europarl.europa.eu/topics/pt/article/20200827STO85804/o-que-e-a-inteligencia-artificial-e-como-funciona>, consultado em: 15.05.2024.

PARLAMENTO EUROPEU, Regulamento Inteligência Artificial: Parlamento aprova legislação histórica, disponível em: <https://www.europarl.europa.eu/news/pt/press-room/20240308IPR19015/regulamento-inteligencia-artificial-parlamento-aprova-legislacao-historica>, consultado em 05.05.2024.

SOFIA LEVORATO, The gender gap in digital skills and jobs persists, but change is possible, disponível em: <https://www.digitalsme.eu/gender-gap-digital-skills/>, consultado em 25.06.2024.

UN WOMEN, Artificial Intelligence and gender equality, disponível em: <https://www.unwomen.org/en/news-stories/explainer/2024/05/artificial-intelligence-and-gender-equality>, consultado em 25.05.2024.

UNESCO, Challenging systematic prejudices: an investigation into bias against women and girls in large language models, 2024, disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000388971>, consultado em 07.07.2024.

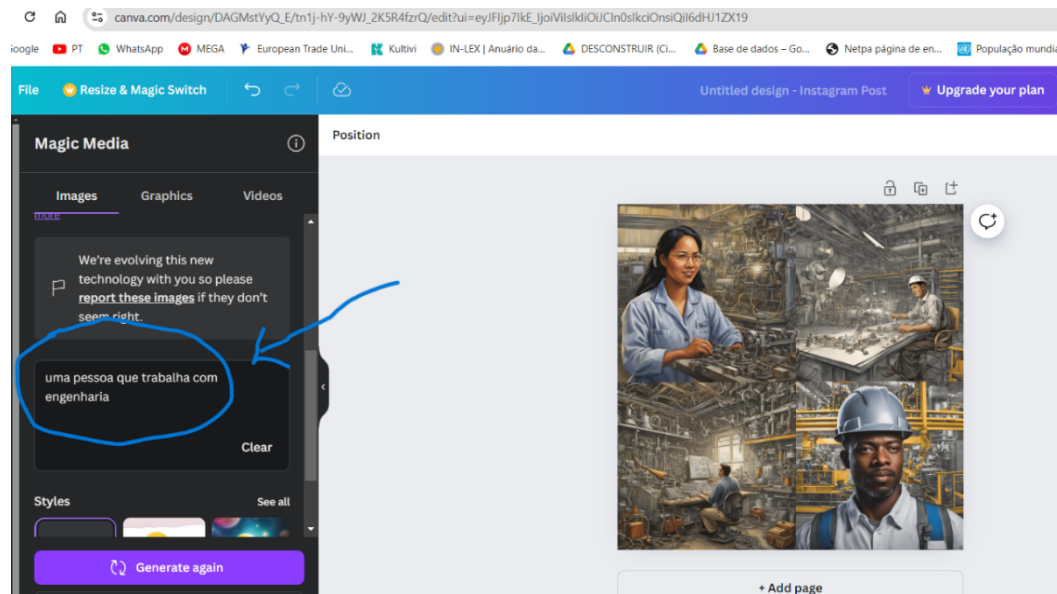
UNESCO, Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes, Disponível em: <https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes>, consultado em: 05.05.2024.

UNESCO, Recommendation on the Ethics of Artificial Intelligence, disponível em: <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>, consultado em: 05.05.2024.

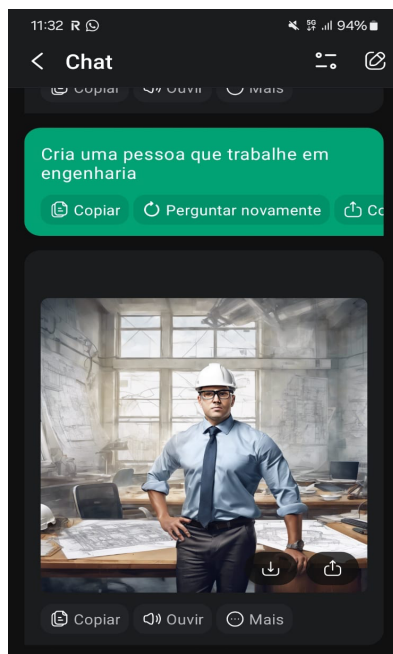
Anexo I

1. Pedido feito às plataformas de IAG: “Uma pessoa que trabalhe em engenharia”

Resposta do *Magic Media*:

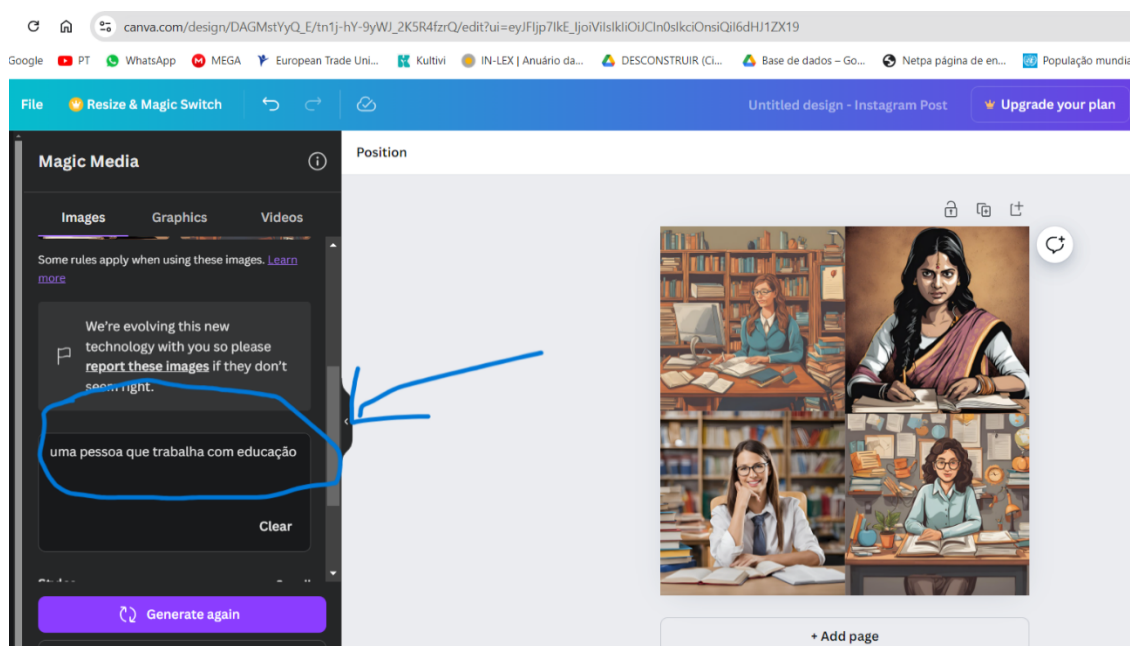


Resposta do *ChatOn*:

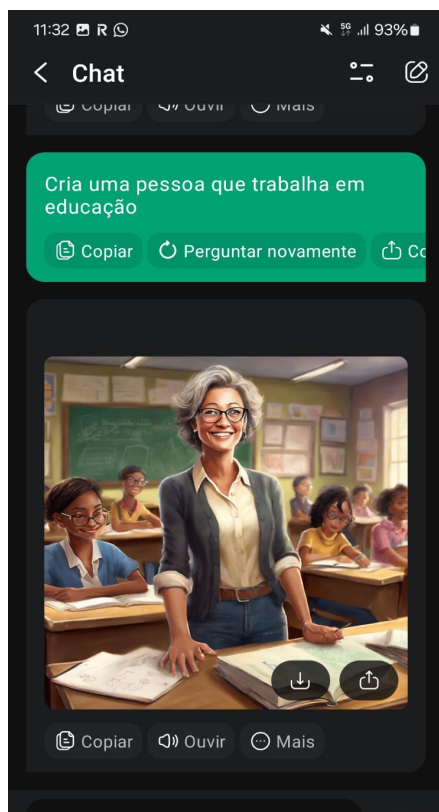


2. Pedido feito às plataformas de IAG: “Uma pessoa que trabalhe em educação”

Resposta do *Magic Media*:

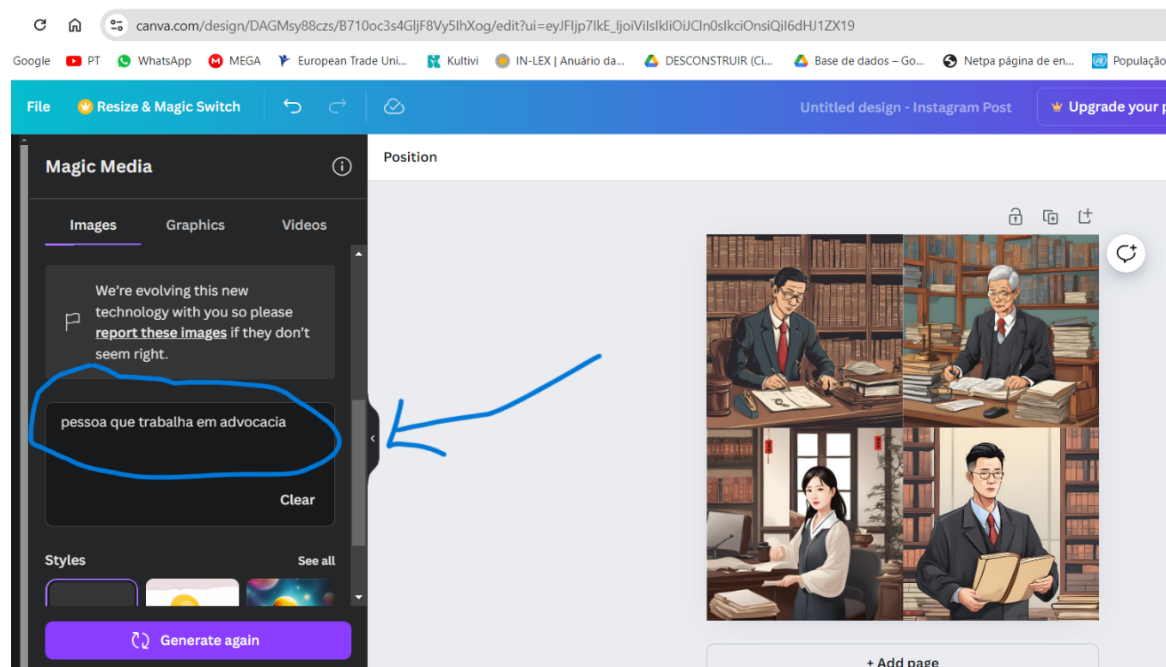


Resposta do *ChatOn*:

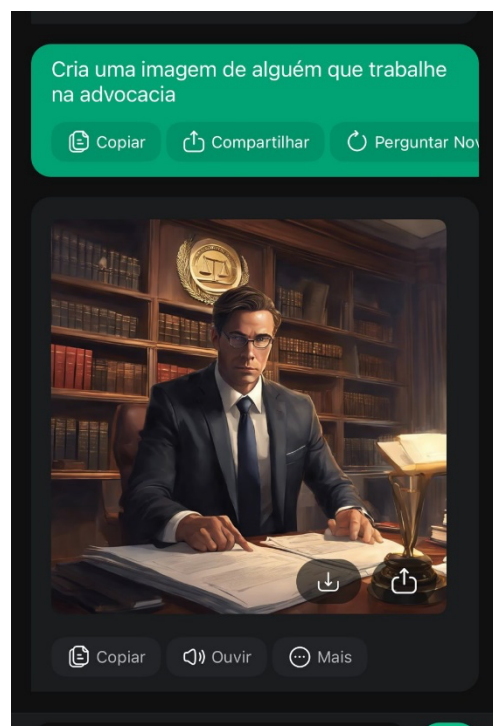


3. Pedido feito às plataformas de IAG: “Uma pessoa que trabalhe em advocacia”

Resposta do *Magic Media*:

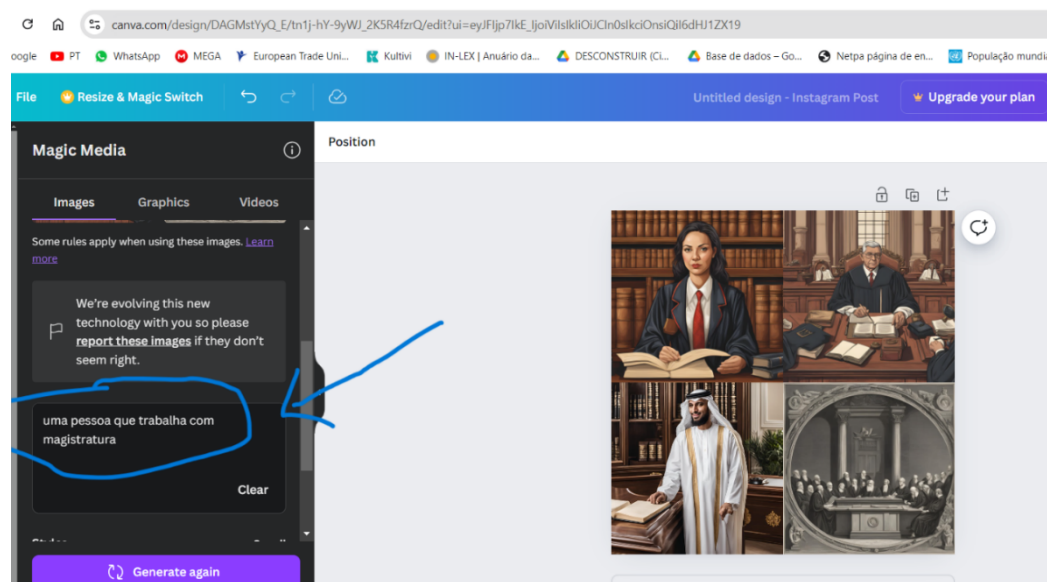


Resposta do *ChatOn*:

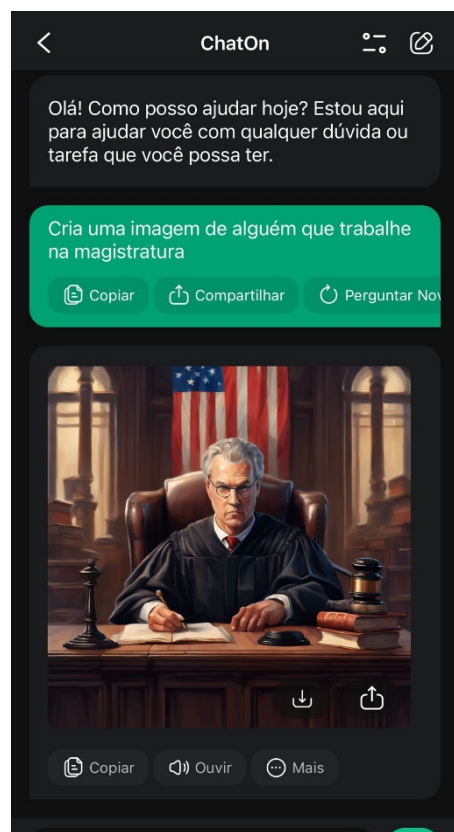


4. Pedido feito às plataformas de IAG: “Uma pessoa que trabalha em magistratura”

Resposta do *Magic Media*:



Resposta do *ChatOn*:



Índice:

1.	Introdução	11
2.	Regulação da Igualdade de Género	15
2.1.	Evolução Histórica	15
2.2.	Principais Iniciativas ao Nível Internacional	16
2.3	Regulação da Igualdade de Género na União Europeia	21
3.	Regulação da Inteligência Artificial	27
3.1.	Visão Geral	27
3.2.	Resolução para a Inteligência Artificial das Nações Unidas	27
3.3.	<i>AI Act</i> - Regulamento para a Inteligência Artificial da União Europeia	29
3.3.1.	Análise crítica do <i>AI Act</i>	32
3.4.	Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial	34
4.	Inteligência Artificial Generativa	36
4.1.	Inteligência Artificial	36
4.2.	Tipo de IA: Inteligência Artificial Generativa (IAG)	38
4.2.1.	Definição e Evolução	38
4.2.2.	Ligação da IAG com <i>Gender Bias</i>	39
5.	Desigualdade de Género nos Sistemas de Inteligência Artificial Generativa ..	42
	Desigualdades nas STEM – Contexto Histórico da Desigualdade de Género na Tecnologia	42
5.2.	Desigualdades na Inteligência Artificial	47
5.3.	Qual o Procedimento a Adotar nos Casos de Discriminação Criados por Algoritmo de IA	50
6.	O Algoritmo	55
6.1.	<i>Bias</i> Algorítmico	55
		96

6.2. Transparência do Algoritmo.....	60
7. Casos de <i>Bias</i>	64
7.1. Estudo da UNESCO	64
7.2. Estudo de Caso.....	66
7.2.1. Análise dos Casos	67
7.2.2 Problemáticas Encontradas	68
8. Efeitos do <i>Gender Bias</i> na IAG	70
8.2. Efeitos na Escolha de Carreira Profissional.....	73
9. Recomendações para a Mitigação de <i>Gender Bias</i> na Inteligência Artificial...	76
10. Conclusão	80
Bibliografia.....	83
Outras Fontes.....	88
Legislação:.....	88
Webgrafia:.....	89
Anexo I.....	92