UNIVERSIDADE CATÓLICA PORTUGUESA

Admissibilidade processual da prova digital:

Os perigos provenientes da Inteligência Artificial Generativa

Francisco Calejo Martins Ribeiro Durães

Mestrado em Direito

Faculdade de Direito | Escola do Porto

2025

UNIVERSIDADE CATÓLICA PORTUGUESA

Admissibilidade processual da prova digital:

Os perigos provenientes da Inteligência Artificial Generativa

Francisco Calejo Martins Ribeiro Durães

Orientador: Pedro Miguel Freitas

Mestrado em Direito

Faculdade de Direito | Escola do Porto

2025

Agradecimentos

A realização desta dissertação representa o culminar de uma etapa académica que só foi possível graças ao apoio, incentivo e orientação de várias pessoas, às quais deixo o meu mais sincero agradecimento.

Em primeiro lugar, um agradecimento especial à minha família, em particular aos meus pais, aos meus avós e à minha irmã pelo apoio incondicional, paciência e compreensão ao longo deste desafio.

Aos meus colegas e amigos, que comigo partilharam esta fase académica e com quem pude desabafar e pedir aconselhamento.

Agradeço também à Universidade Católica Portuguesa pelo seu ambiente de excelência académica e aos docentes do Mestrado em Direito que, com os seus contributos e partilhas, enriqueceram o meu percurso académico e pessoal.

Ao meu orientador, o Exmo. Professor Doutor Pedro Freitas pela sua constante disponibilidade, orientação rigorosa e apoio científico ao longo de todo o processo de escrita.

Por fim, agradeço a todas as pessoas que, direta ou indiretamente, contribuíram para a concretização deste marco.

A todos, o meu reconhecimento e gratidão.

Resumo

A seguinte dissertação analisa os desafios jurídicos colocados pela Inteligência Artificial Generativa no contexto da prova digital no processo penal, com particular destaque para plataformas como o *ChatGPT*, que levantam sérias questões quanto à autenticidade da prova e à equidade processual.

Não obstante a IA Generativa trazer inegáveis benefícios — como o aumento da produtividade e o apoio na redação jurídica — a mesma representa riscos significativos para os princípios fundamentais do direito processual penal, como a busca da verdade material e o direito a um julgamento justo.

É ainda abordado o uso da IA em tecnologias de justiça preditiva e reconhecimento facial, que evidenciam preconceitos raciais e desigualdades socioeconómicas. Especial atenção é dada ao impacto dos *deepfakes*, capazes de manipular fotografias, vídeos e áudios com elevado realismo, corrompendo o valor probatório e exigindo uma perícia técnica sofisticada para a sua verificação. É ainda questionada a atual eficácia das normas processuais penais, defendendo-se a urgência de uma regulamentação adaptada à era digital.

Neste trabalho, propomos soluções como o reforço da literacia digital, a democratização do acesso à perícia forense e a criação de mecanismos legais específicos para garantir a integridade da prova. Sem uma resposta jurídica robusta e equitativa, o uso da IA pode comprometer a justiça penal, especialmente para os mais vulneráveis.

Palavras-chave: Inteligência Artificial, Inteligência Artificial Generativa, *ChatGPT*, *deepfakes*, Processo Penal, prova, perícia.

Abstract

The following dissertation analyses the legal challenges posed by Generative Artificial Intelligence in the context of digital evidence in criminal proceedings, with particular emphasis on platforms such as *ChatGPT*, which raise serious questions about the authenticity of evidence and procedural fairness.

Although Generative AI brings undeniable benefits - such as increased productivity and support in legal writing - it poses significant risks to the fundamental principles of criminal procedural law, such as the search for material truth and the right to a fair trial.

The use of AI in predictive justice and facial recognition technologies, which highlight racial prejudice and socio-economic inequalities, is also addressed. Special attention is paid to the impact of deepfakes, which are capable of manipulating photographs, videos and audio with high realism, corrupting their probative value and requiring sophisticated technical expertise to verify. The current effectiveness of criminal procedural rules is also questioned, and the urgent need for regulations adapted to the digital age is defended.

In this paper, we propose solutions such as strengthening digital literacy, democratising access to forensic expertise and creating specific legal mechanisms to guarantee the integrity of evidence. Without a robust and equitable legal response, the use of AI can jeopardise criminal justice, especially for the most vulnerable.

Keywords: Artificial Intelligence, Generative Artificial Intelligence, *ChatGPT*, *deepfakes*, Criminal Procedure, evidence, expertise.

Índice

Abreviaturas e Siglas

ANN	*Artificial Neural Network*
ART	Artigo
CAAD	Estrutura do Centro de Arbitragem Administrativa
CAC	Administração do Ciberespaço da China
CCTV	Câmaras de Circuito Fechado de Televisão
CEPEJ	Comissão Europeia para a Eficiência da Justiça
CNN	*Convolutional Neural Network*
COMPAS	*Correctional Offender Management Profiling for Alternative Sanctions*
CPP	Código de Processo Penal
DNN	*Deep Neural Networks*
EU AI Act	*Ato de Inteligência Artificial da União Europeia*
F1	Fórmula 1
HART	*Human Assisted Review Tool*
IA	Inteligência Artificial
iOMS	*Innovative Prison Systems*
LGPD	Lei Geral de Proteção de Dados Pessoais
LLM	*Large Language Models*
ONG	Organização Não Governamental
PGR	Procuradoria-Geral da República
PIPL	Lei de Proteção de Informações Pessoais
RGPD	Regulamento Geral de Proteção de Dados
TSE	Tribunal Superior Eleitoral

1. Introdução

A transformação digital das sociedades modernas tem vindo a provocar profundas alterações no modo como a justiça é administrada, colocando novos desafios ao Direito, em especial na esfera penal. Entre as inovações mais marcantes está a ascensão da Inteligência Artificial Generativa, cujas ferramentas — como o *ChatGPT*, o *DALL·E* ou os *deepfakes* — revelam um enorme potencial de utilidade, mas também riscos consideráveis para os princípios fundamentais do processo penal, nomeadamente a busca pela verdade material, a livre apreciação da prova e o direito a um julgamento justo.

Esta dissertação analisa o impacto crescente da IA Generativa na produção e admissibilidade de prova digital em sede de processo penal, destacando os riscos associados à sua utilização em casos que envolvam arguidos particularmente vulneráveis, ou cidadãos provenientes de meios socioeconómicos mais desfavoráveis.

A utilização de algoritmos opacos, a dificuldade em verificar a autenticidade de conteúdos audiovisuais e a possibilidade de decisões judiciais serem influenciadas por sistemas automatizados levantam sérias dúvidas quanto à equidade processual, à transparência e à segurança jurídica. Casos emblemáticos e estudos recentes demonstram como a IA pode não só reproduzir preconceitos e desigualdades já existentes, mas também introduzir um novo grau de complexidade técnica que limita o contraditório e o exercício pleno do direito de defesa.

Ao longo deste trabalho, será feita uma abordagem crítica a estes fenómenos, refletindo sobre o seu enquadramento jurídico, os mecanismos legais de controlo e a necessidade de adaptação dos princípios processuais à realidade digital. Serão ainda analisadas propostas regulatórias, como o EU AI Act, bem como possíveis soluções para garantir a integridade da prova digital, o acesso equitativo à perícia técnica e a salvaguarda dos direitos fundamentais no contexto de um processo penal cada vez mais alvo da tecnologia.

2. Inteligência Artificial: conceitos, riscos e benefícios

a. Inteligência Artificial: a 4ª Revolução Industrial

A inteligência está profundamente conectada às atividades humanas, sendo essencial para diversas tarefas, como a compreensão de línguas, a expressão e a aprendizagem. Diferentes formas e níveis de inteligência manifestam-se tanto em seres humanos como em outras espécies.

Porém, uma nova era está a surgir, reconhecida como a "4ª Revolução Industrial, marcada por vários desafios como consequência da convergência de tecnologias digitais, físicas e biológicas"[1].

A Inteligência Artificial (IA) está a ter um impacto sem precedentes e os avanços neste domínio que presentemente assistimos revelam-se revolucionários e abrem portas a um novo horizonte capaz de mudar o rumo da civilização humana. É inquestionavelmente uma ferramenta poderosa, mas, como qualquer ferramenta, para utilizá-la da maneira mais eficaz e eficiente, é essencial compreender não apenas os seus inúmeros benefícios, mas também os seus riscos, limitações e possíveis ameaças[2].

De acordo com a investigação realizada "Atitudes face ao impacto da digitalização e automatização sobre a vida quotidiana": "61% dos europeus possuem uma opinião positiva acerca da inteligência artificial, mas 88% consideram que estas tecnologias exigem uma gestão com cautela (Eurobarómetro 2017, UE28)"[3].

Atentemos à noção de IA, que é apresentada por vários autores.

John McCarthy (2007) define Inteligência Artificial (IA) como "a ciência e engenharia de criar máquinas inteligentes, especialmente programas de computador inteligentes. Ela está relacionada à tarefa de usar computadores para entender a inteligência humana, mas a Inteligência Artificial não precisa de se limitar a métodos observáveis biologicamente". Esta definição destaca a IA como uma ciência e uma engenharia com características próprias, sugerindo que a IA não se deve restringir a imitar processos naturais. É possível inferir a ideia de que a IA deve abrir espaço para a imaginação, a arte e a ficção, assim como os aviões só foram capazes de voar ao abandonar a tentativa de replicar o bater das asas dos pássaros, seguindo a imaginação humana.

De forma mais simples, Luger e Stubblefield (1998) apresentam a IA como uma disciplina voltada ao estudo e criação de entidades artificiais com habilidades cognitivas similares às humanas (Costa & Simões, 2008). A IA é, portanto, tanto uma ciência, que busca compreender o fenômeno da inteligência, quanto uma área da engenharia, dedicada a criar ferramentas que auxiliem o ser humano (Russel & Norvig, 1995). Sistemas Inteligentes têm o propósito de capacitar computadores a executar tarefas que dependem do

[1] Novais & Freitas, 2018.
[2] Villasenor, 2024.

[3] Rocha, 2023.

conhecimento e do raciocínio humano. A habilidade humana de agir inteligentemente está frequentemente ligada ao conhecimento, que é essencial na construção desses sistemas[4].

Resumindo, A IA é um conceito amplo que engloba diversos métodos computacionais projetados para executar atividades que, geralmente, exigiriam a capacidade cognitiva humana — como entender a linguagem, identificar padrões, tomar decisões e aprender com a prática. É tida como discriminativa, isto é, procura identificar padrões nos dados para realizar classificações e previsões.

Tendo bem presente esta ideia, cumpre esclarecer a diferença entre a Inteligência Artificial e a Inteligência Artificial Generativa dado que a linha que separa estes dois conceitos parece ser algo ténue.

A IA Generativa é uma subcategoria da IA, baseada em modelos generativos profundos, capazes de criar conteúdos realistas, como textos, imagens, código ou sons. a partir de dados existentes. Utiliza uma tecnologia de base denominada d*eep learning,* sendo que a sua característica mais marcante parece ser a sua versatilidade preparada para criar conteúdo original – o mesmo input pode gerar resultados diferentes, pois a geração é probabilística e não determinista[5].

É inegável que a Inteligência Artificial tem sido um domínio alvo de um interesse académico e profissional desde meados do século XX. Por outro lado, a Inteligência Artificial Generativa apenas recentemente começou a atrair um vasto interesse público, principalmente com o lançamento a 30 de novembro de 2022 do ChatGPT[6] pela *OpenAI*, empresa fundada no Estado da Califórnia nos EUA, considerado pela *Harvard Business Review* como "um ponto de viragem para a Inteligência Artificial"[7].

Definido pela *OpenAI*, o *ChatGPT* utiliza um "algoritmo de aprendizagem automática (*machine learning*) autónomo, utilizando LLMs (*large language models*), baseada em redes neurais e, mais especificamente, artificiais (ANNs), treinados para analisar o texto na Internet, desenvolvendo um modelo estatístico que lhe permite juntar palavras em resposta a um determinado pedido"[8]. Assim, apresenta-se como um *chatbot* genuinamente útil para qualquer tarefa, seja ela a resposta longa a pedidos ou questões, a criação de um *software* ou a simples redação de um plano de perda de peso.

A 15 de março de 2023, foi lançado o *ChatGPT-4*, uma atualização do modelo usado para o *ChatGPT*, que, ao "invés de processar somente textos, gera ainda imagens e textos como *inputs* e conta com uma maior precisão de atuação". Além desta plataforma, a empresa também lançou o *Dall-e2*, um sistema gerador de imagens e a Whisper, um "sistema automático de reconhecimento que faz a conversão de áudio em texto e a tradução para

[4] Ibidem em 1.
[5] Banh, 2023.
[6] *OpenAI*, 2022. *Introducing ChatGPT.* Disponível em https://openai.com/index/chatgpt/.
[7] Mollick, 2022.
[8] Zahn, 2022.

diversos idiomas"[9], algo que teremos a oportunidade de analisar cuidadosamente no próximo capítulo.

Até ao presente momento, o modelo mais eficiente e mais recente da *OpenAI* é o *ChatGPT-4.1*, lançado a 14 de abril de 2025, um modelo especializado em tarefas de código. A empresa detentora deste sistema aponta-o com um nível de precisão impressionante de 55%. Os especialistas destacam ainda a capacidade do modelo de seguir instruções particularmente complexas no desenvolvimento de apps funcionais a partir de *prompts*, ou seja, indicações que podemos fornecer a esta plataforma para que a mesma possa gerar respostas ou realizar tarefas específicas[10].

Esta capacidade infindável de resposta do *ChatGPT* não passou despercebida. Pelo contrário, gerou uma enorme onda de surpresa dado que a utilização da Inteligência Artificial se resumia a domínios mais simples como a pesquisa na Internet. É inequívoca esta transição proporcionada pelo *ChatGPT*, que é acessível a todos.

b. Riscos associados à Inteligência Artificial

Com toda esta acessibilidade proporcionada pela IA, surgem várias questões em torno dos seus riscos. Uma delas (e a mais controversa), sendo alvo de enorme debate entre os entendidos, é a seguinte: poderão as máquinas substituir os humanos?

Apesar de não haver certezas do estatuto que estas plataformas de IA terão a longo prazo, é evidente que alguns empregos serão transformados, alguns poderão desaparecer, e outros irão surgir, ou, pelo menos, a sua relação de trabalho poderá vir a ser alterada com a redução da carga horária dos trabalhadores. Também é de esperar que seja cada vez mais procurada mão de obra especializada, isto é, trabalhadores dotados de literacia digital que possam programar e alimentar as máquinas.

Além deste risco de redução dos esforços humanos, com a IA Generativa, a diferença na distinção entre a verdade e a mentira parece cada vez mais acentuada e as chamadas *fake news* são cada vez mais frequentes. Exemplo disso foi uma entrevista com Michael Schumacher divulgada pela revista alemã *Die Aktuelle,* na qual publicaram uma foto do célebre piloto da F1 na capa com o título "Michael Schumacher, the first interview"[11], após o mesmo não ter sido visto em público desde o acidente que sofreu no cérebro enquanto esquiava em 2013. E só no final do artigo é que era possível confirmar que a entrevista tinha sido criada por IA.

Assim, estes modelos de IA podem também atuar com o objetivo de persuadir e influenciar a opinião pública, degradando a confiança da sociedade, sendo necessário uma abordagem conjunta entre meios de comunicação social e programadores de modo a identificar produtos gerados pela IA.

---

[9] Rossetti, 2023.
[10] Parreira, 2025.
[11] Mncwabe & Schmidt, 2023.

No mesmo sentido, a IA também alberga outras legítimas preocupações em termos de ética. Sabemos que o desenvolvimento de novas tecnologias deve acompanhar princípios éticos, como a não discriminação, a equidade, a responsabilidade, a privacidade e a proteção de dados e a transparência, além de interesses individuais, coletivos e comerciais.

Recentemente, surgiu uma notícia alarmante de que a plataforma de IA da *Google* (o *Gemini*), concorrente do *ChatGPT*, trocou mensagens agressivas e pejorativas com uma estudante universitária norte-americana que precisava de ajuda para um trabalho académico sobre "o envelhecimento da população e os desafios que as tendências demográficas colocam para as sociedades"[12]. Este chatbot dirigiu-se à jovem, apelidando-a de "fardo para a sociedade", "nódoa no universo" e "por favor, morre".

Este incidente destaca as limitações associadas ao uso de IA em contextos mais sensíveis, convocando problemas relacionados com a responsabilidade das empresas e riscos psicológicos. As empresas que desenvolvem este tipo de plataformas têm a responsabilidade de garantir que as mesmas são seguras e confiáveis e que não há perigo de dar respostas imprevisíveis e inadequadas como a supramencionada. Estas interações podem ter consequências bem reais conforme bem referiu a estudante já que alguém mais vulnerável ou com a sua saúde mental debilitada colocar-se-ia numa situação extrema. Afinal, as ferramentas de IA não possuem consciência e sensibilidade, sendo antes um produto dos humanos, reproduzindo os seus valores. É um alerta de ainda um longo caminho a percorrer no desenvolvimento destas plataformas.

De modo a concluir, a IA Generativa é uma tecnologia emergente extremamente preponderante na redação e criação de textos, não só no domínio jurídico – esta vertente será tratada no último capítulo - como em tantos outros. Naturalmente que aqueles que aprenderem e forem capazes de utilizar as ferramentas de IA poderão trabalhar de forma mais eficaz, poupando tempo e recursos. Por muito poderosa que seja a IA, apesar de algumas funções poderem vir a ser substituídas (como funções de tradução e atendimento ao cliente), há outras que a mesma não irá substituir no futuro – no caso de um advogado ou de um magistrado, o aconselhamento e o desenvolvimento de relações genuínas com clientes ou colegas de profissão ou a simples orientação a estagiários são tarefas que nenhuma plataforma de IA poderá fazer.

Porém, esta preponderância de sistemas de IA Generativa está ainda aliada à criação de imagens e vídeos que já estão a ser apresentados em sede audiência de julgamento. Daqui decorrem sérias preocupações e efeitos nefastos para a veracidade processual, afetando Direitos Fundamentais dos cidadãos e a boa aplicação da justiça, como veremos adiante.

É comum caracterizarmos uma norma jurídica como rígida e imutável. Contudo, é essencial criar normas que regulem as relações jurídicas decorrentes da sociedade na evolução tecnológica em que vivemos, considerando duas frentes complementares: por um lado, estabelecer um quadro jurídico geral que sirva de base para toda a legislação

[12] Caetano, 2024.

sobre inteligência artificial; por outro, aprovar leis específicas adaptadas a diferentes domínios da IA, capazes de responder a desafios concretos de cada área[13].

c. Impacto da Inteligência Artificial na Justiça Penal

Em relação mais concretamente à Justiça Penal, é agora o momento oportuno para problematizar o impacto que a IA pode ter na mesma e tratar da questões como os algoritmos, aas câmaras de reconhecimento facil e a justiça preditiva, cuja permissa, segundo Godefroy, Lebaron e Lévy-Vehel, baseia-se na utilização de dados passados, métodos estatísticos e técnicas de *machine learning* automática para estimar a probabilidade de eventos futuros. No entanto, esse processo levanta preocupações éticas, especialmente no que diz respeito à previsão de decisões[14]. Neste sentido, podemos então sugerir que a Inteligência Artificial dá lugar a uma forma de "justiça preditiva", ao criar perfis pessoais que podem ser discriminatórios.

A aplicação da IA em programas utilizados no ramo do Direito Penal, recorrendo a algoritmos para delinear o perfil de pessoas potencialmente criminosas, pode implicar o "efeito de "círculos viciosos" e "profecias autorrealizadas": as vizinhanças consideradas em risco atraem mais atenção da polícia e a polícia deteta mais crime, o que leva a uma vigilância policial excessiva das comunidades que nelas vivem (…) [dando origem a] uma possível "tirania do algoritmo" que poderia minimizar ou mesmo substituir progressivamente o julgamento humano, conforme refere a Comissão Europeia para a Eficiência da Justiça"[15].

Dois exemplos desta justiça preditiva são os *softwares* COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) e HART (*Human Assisted Review Tool*).

O primeiro teve origem nos Estados Unidos da América, criado pela empresa *Northpointe*, usado em alguns tribunais federais como em New York, Wisconsin ou California e que tem como objetivo auxiliar a decisão judicial, avaliando o risco de reincidência de um determinado indivíduo. Para tal, são tidas em conta várias escalas, tais como uma ferramenta de avaliação de risco pré-julgamento que se liga à probabilidade de uma pessoa não comparecer em tribunal; uma escala de reincidência geral, concebida para prever a probabilidade de um indivíduo cometer novos atos criminosos após a sua libertação; e uma escala de reincidência violenta que mede a probabilidade de esse indivíduo voltar a cometer infrações graves e violentas depois de uma condenação anterior[16]. Dizer ainda que este programa incluiu também 137 perguntas respondidas pelo réu, ou informações extraídas de registos criminais, nomeadamente, o seu histórico familiar e histórico criminal, entre outros, em que o algoritmo classifica a pessoa numa

[13] Sepúlveda & Machuca, 2024.
[14] Godefroy, Lebaron e Lévy-Vehel, J., 2019.
[15] CEPEJ, 2018.
[16] Deb, 2023.

escala de 1 (baixo risco) a 10 (alto risco), sendo que as suas conclusões auxiliam o juiz a decidir sobre a sentença.

Porém, segundo um estudo realizado pela ONG *ProPublica*, o programa avaliativo apresentou um viés, atribuindo incorretamente um risco mais elevado de reincidência criminal a indivíduos afro-americanos, sendo essa taxa identificada como duas vezes superior à atribuída a pessoas de outras origens[17]. Assim, é possível apontar pelo menos duas causas para esta discriminação racial: por um lado, o preconceito pessoal do programador, que pode ser refletido ao longo do desenvolvimento do algoritmo, embora disso não se apercebendo; por outro lado, os múltiplos fatores considerados pelo algoritmo na avaliação do réu, que podem refletir desigualdades já existentes. Exemplificando, além de questões raciais, indivíduos instruídos teriam pontuações de risco relativamente mais baixas no COMPAS do que indivíduos sem escolaridade; réus com trabalho a tempo inteiro e salários mais elevados receberiam pontuações de risco mais baixas do que aqueles que se encontram desempregados. Com estes pressupsotos, não será difícil concluir que alguns algoritmos de IA e softwares a ela inerentes sejam tendenciosos contra certos grupos e minorias, destacando a sua fragilidade social e económica, estando a racionalidade, a equidade e a justiça postas de lado.

Já o programa HART vem sido a ser testado desde 2017 e o software foi programado tendo por base decisões policiais de arquivos da Polícia de Durham de 2008 a 2012. A partir dessa informação, também ele classifica o risco de reincidência do respetivo indivíduo, como sendo baixo, moderado ou alto. A maioria desses elementos está relacionada com o histórico criminal do indivíduo, sendo complementada por informações como a quantidade de relatórios policiais associados a ele, além de outras variáveis, como a idade, o género e dois tipos de códigos postais referentes à sua residência[18].

Esse último fator é particularmente significativo, pois uma investigação realizada em 2020 pela organização do Reino Unido *Big Brother Watch* — defensora das liberdades civis e dos direitos humanos — revelou que uma das variáveis dos códigos postais utilizados pelo sistema HART tinha origem num produto de dados de marketing da empresa global de análise de dados *Experian*, denominado *Mosaic*. O *Mosaic* é uma "ferramenta de segmentação social, geográfica e demográfica" que classifica determinadas áreas, utilizando estereótipos associados aos códigos postais. Essa classificação é construída a partir de um vasto conjunto de cerca de 850 milhões de dados, incluindo informações censitárias, origem étnicas, dados de saúde, situação laboral, benefícios sociais, nomes próprios e de família associados a grupos étnicos, bem como dados recolhidos de fontes *online*, entre outros. Estas informações são utilizadas para perfilar cerca de 50 milhões de adultos no Reino Unido, com base nos seus códigos

[17] Ibidem em 16.
[18] Pedroso & Santos, 2024.

postais, gerando perfis estereotipados ao associar determinadas características demográficas a cada grupo ou área[19].

Os métodos adotados pelo programa HART (e pelo COMPAS) foram amplamente criticados por se basearem numa análise estatística de uma amostra de dados que acaba por gerar discriminação negativa contra determinados indivíduos e grupos sociais[20].

Como bem consta na Comissão Europeia para a Eficiência da Justiça, a verdade é que "existem riscos potenciais de discriminação quando se considera que estes instrumentos, que são construídos e interpretados por seres humanos, podem reproduzir desigualdades injustificadas e já existentes no sistema de justiça penal em causa, em vez de corrigir certas políticas problemáticas"[21] e a tecnologia pode acabar por validá-las.

Cumpre apenas mencionar que na Europa Continental ainda é praticamente inexistente a utilização de instrumentos preditivos. Tal poderá dever-se ao modelo de justiça penal dos sistemas anglo-saxónicos, que influencia os objetivos da punição e atribui particular importância à fase de julgamento — momento central para a análise da culpa — onde a atuação humana permanece indispensável. Além disso, é vedado o uso exclusivo do tratamento automatizado de dados pessoais para fundamentar decisões ou para outras finalidades no contexto do processo penal, com base na Lei n.º 58/2019, de 8 de agosto, que assegura a execução do RGPD[22].

No ordenamento jurídico português, mesmo que ainda não existam casos concretos, no que se refere ao uso da inteligência artificial no apoio às decisões judiciais, a posição do poder judicial é clara: as tecnologias de IA não devem substituir o papel do juiz na tomada da decisão, nem exercer qualquer influência negativa ou tendenciosa sobre a fundamentação das sentenças[23].

No entanto, pode referir-se como exemplo um projeto piloto desenvolvido pela empresa portuguesa IPS — *Innovative Prison Systems* — que utiliza IA e análise preditiva na gestão penitenciária. Trata-se do software "*Horus 360º iOMS*" (*Intelligent Offender Management System*), concebido para apoiar decisões no âmbito da justiça criminal. Este sistema inteligente acompanha todo o percurso da pessoa na justiça — desde a detenção, passando pelo julgamento, até à prisão e reintegração — propondo medidas com base no seu perfil individual. Entre as suas funções estão a gestão de detenção, decisões judiciais, programas de reabilitação e libertação antecipada. Embora ainda não esteja em uso em Portugal, está em fase de testes em vários países europeus, despertando o interesse de diversas instituições ligadas à justiça, como os órgãos de Polícia Criminal, Tribunais e serviços prisionais, exigindo, no entanto, adaptações às especificidades legais de cada país[24].

[19] Big Brother Watch, 2020.
[20] Reiling, 2020.
[21] Ibidem em 15.
[22] Anabela Rodrigues, 2023.
[23] Lameira, 2021.
[24] Agência Lusa, 2023.

No que toca aos algoritmos complexos, como os usados nestes sistemas preditivos ou de reconhecimento biométrico, estes exigem conhecimentos especializados para serem questionados. Se até advogados experientes enfrentam dificuldades para contestar resultados algorítmicos devido à falta de acesso aos dados de treino e à lógica interna dos sistemas, para inimputáveis, por exemplo, que raramente têm recursos para contratar peritos independentes, essa barreira é ainda mais proeminente, como analisaremos adiante.

Um acórdão do Tribunal da Relação de Lisboa foi considerado suspeito de ter sido influenciado pela IA devido a citações de leis inexistentes, ilustra a dificuldade de identificar e contestar erros algorítmicos[25]. O Conselho Superior da Magistratura recusou-se a investigar enquanto o processo estivesse em curso deixando os arguidos (imaginemos que se tratassem daqueles especialmente os mais vulneráveis, como os inimputáveis) sem meios imediatos para questionar a validade das provas.

Com base na revisão da literatura, os desafios relacionados ao uso da IA na justiça decorrem principalmente das escolhas feitas na seleção e na programação dos algoritmos utilizados. "Um dos principais riscos diz respeito à transparência processual, relacionada com a opacidade dos algoritmos. A programação dos algoritmos não permite compreender como os *inputs* chegaram aos resultados, gerando um efeito de caixa-preta ("*black box*") "[26].

A Associação Sindical dos Juízes Portugueses alertou para os riscos de sistemas automatizados que carecem de explicabilidade, alertando que o "automatismo aumenta a probabilidade de erro"[27]. Para inimputáveis, cuja participação no processo já é limitada por sua condição, essa opacidade é particularmente prejudicial e contribuiu para a desvirtuação e a perda das características humanas das decisões judiciais. No contexto jurídico, é fundamental que as novas tecnologias utilizadas para auxiliar nas decisões judiciais atuem como ferramentas de suporte ao juiz, apresentando a fundamentação exigida pela lei.

Além disso, os algoritmos de IA frequentemente refletem vieses presentes nos dados com os quais foram treinados, o que pode levar a resultados discriminatórios[28]. A IA, por depender da intervenção humana, reflete as escolhas éticas de quem a desenvolve e a alimenta com dados, os quais servem de base para suas decisões. Com o uso da aprendizagem da máquina, esses sistemas conseguem adaptar-se a novas situações e identificar padrões além dos inicialmente previstos, o que faz com que conflitos de interpretação e dificuldades já existentes sejam apenas replicados noutro formato[29]. Por outras palavras, se os dados utilizados forem parciais ou enviesados, a IA tende a reforçar e até ampliar as desigualdades sociais e os preconceitos neles presentes.

[25] Henriques, 2024.
[26] Ibidem em 18.
[27] Agência Lusa, 2023.
[28] Ibidem em 18.
[29] Martins, 2021.

O uso de dados pela IA traz então o risco de as decisões aparentarem serem imparciais e objetivas, quando na verdade não são. Tal ameaça a autonomia do juiz, compromete a sua independência e pode colocar em perigo a garantia de um julgamento justo, transformando-se, assim, num problema de ordem institucional e constitucional. Na prática, "a interpretação do direito processual acaba por influenciada por aqueles que projetam, desenvolvem ou controlam o software, que, ao utilizar correlações estatísticas entre dados, presta apoio ao juiz nas suas decisões"[30]. Este processo pode ocultar ou disfarçar a verdadeira dimensão da influência exercida sobre o julgador, tornando-se potencialmente enganoso.

Também o reconhecimento facial integrado em Câmaras de Circuito Fechado de Televisão (CCTV) utiliza IA para identificar ou verificar a identidade de uma pessoa a partir de imagens ou vídeos capturados por câmaras, com uma capacidade de processamento de 32 segundos. Tal processo envolve a captura de imagem, registando o rosto da pessoa, a análise de características através da avaliação de pontos faciais para criar um modelo digital único, através de algoritmos baseados em *deep learning*, e a comparação com bancos de dados de rostos já armazenados. Esta tecnologia é particularmente usada para segurança pública e partilha dos mesmos riscos que os sistemas acima descritos – precisão, vieses e uso indevido[31].

São imensas as preocupações quanto à segurança e à fiabilidade dos sistemas. A responsabilidade — ou "*accountability"* — por estas tecnologias destaca a necessidade de cumprir rigorosamente as normas e a legislação aplicável, especialmente no que diz respeito à proteção da privacidade e dos dados pessoais, à definição da titularidade dos algoritmos, bem como à salvaguarda da segurança da informação e da integridade dos dados e dos próprios sistemas. A má utilização de dados pessoais e as violações da privacidade tornam este tema particularmente relevante, uma vez que a sua dimensão técnica está intrinsecamente ligada a questões éticas (como a discriminação), à confiança dos utilizadores (como a integridade dos sistemas) e à segurança (como ataques a dados e infraestruturas digitais)[32].

3. Obsolescimento dos atuais princípios basilares do Direito Processual Penal Português

a. Considerações gerais e tipos de deepfakes

Inicialmente, apesar de a Inteligência Artificial Generativa ter revestido a sua maior atenção no que à geração de texto diz respeito, é de extrema importância não menosprezar a possibilidade de este instrumento ser também utilizado para produzir imagens e vídeos.

[30] Contini & Reiling, 2022.
[31] Wall, 2019.
[32] Ettekoven & Prins, 2018.

Em janeiro de 2021, ainda antes da introdução no mercado do *ChatGPT*, a *OpenAI* anunciou o lançamento público da *DALL-E,* apontada pela mesma como "um salto em frente na nossa capacidade de gerar imagens que correspondem exatamente ao texto fornecido"[33]. Uma nova versão, *DALL-3*, foi lançada em outubro de 2023.

É legítimo perguntarmo-nos o que traz de novo a IA na criação e manipulação de imagens e a resposta baseia-se na credibilidade e na acessibilidade. O desenvolvimento de técnicas de *machine learning* e inteligência artificial aplicadas à alteração e modificação de imagem e áudio permitiu que "a criação de resultados praticamente indistinguíveis da realidade, sobretudo na área da fotografia, se tornasse perfeitamente acessível mesmo para quem não tenha profundos conhecimentos técnicos"[34]. Através destas poderosas técnicas, torna-se então possível manipular ou gerar conteúdos visuais e sonoros com um elevado potencial fraudulento.

Este fenómeno, denominado como *deepfakes,* começou quando um utilizador anónimo da plataforma *Reddit*, inseriu rostos de celebridades desconhecidas em vídeos pornográficos. Ao divulgar o código informático responsável pela criação dos *deepfakes* (uma combinação de *'deep learning'* e *'fakes'*), despertou um enorme interesse na comunidade, originando uma proliferação massiva de conteúdos manipulados[35].

Um dos primeiros *deefakes* a ser criado demonstrou verdadeiramente o poder da IA e consistiu no "*Synthesizing Obama*", de 2017. Foi utilizada na perfeição a tecnologia de sincronização labial com base não só em imagens como também de áudios existentes. Hoje, é possível assistirmos a qualquer líder ou figura mundial, seja de que nacionalidade for, a falar ou discursar de forma extremamente coerente e convincente sem nos darmos conta que o mesmo discurso foi proferido por outra pessoa. Tal demonstra novamente o caráter alarmante e a necessidade de adotar medidas tecnológicas, sociais e jurídicas adequadas para tentar suprimir este advento dado que, quer queiramos quer não, cada vez mais imagens, vídeos e áudios serão partilhados.

Primeiramente, cumpre perceber e analisar os diferentes tipos de *deepfakes*, sabendo que os dados de treino para as criar podem ser aplicados de várias formas, seja em vídeo ou imagem.

Segundo a Europol[36], e provavelmente a categoria na qual estamos mais familiarizados, surge a troca de rosto, que consiste na transferência do rosto de uma pessoa para o da pessoa no vídeo; a edição de atributos, alterando caraterísticas da pessoa no vídeo, por exemplo, o estilo ou a cor do cabelo; a reencenação de rosto, que configura a transferência das expressões faciais do rosto de uma pessoa para a pessoa no vídeo alvo; e, por fim, o material totalmente sintético, sendo utilizado material real para treinar o aspeto das pessoas, mas a imagem resultante é totalmente inventada[37].

---

[33] OpenAI, 2023. Disponível em: https://openai.com/index/dall-e-3/.
[34] Freitas, 2023.
[35] Kietzmann & Lee & McCarthy & C. Kietzmann, 2019.
[36] Europol, 2022.
[37] Ver, como exemplo, https://www.thispersondoesnotexist.com/.

Para gerar dados de qualidade, é necessário dispor de uma grande quantidade e variedade de exemplos, que apresentem representações semelhantes, mas com pequenas variações nas mesmas características, para que o processo funcione de maneira consistente[38]. Por exemplo, se uma base de dados contém maioritariamente imagens de homens brancos com cabelo preto, a sua performance não será perfeita na criação de mulheres asiáticas com cabelo louro. À medida que o número e o volume das bases de dados disponíveis crescem, a qualidade e quantidade de dados de treino aumenta. Este facto permitiu que os modelos que criam *deepfakes* aumentem em sofisticação.

Portanto, como o próprio nome sugere, e no mesmo seguimento do que já foi acima mencionado na produção de textos pela IA, o principal componente para a criação de *deepfakes* é o *deep learning*, uma técnica que "manuseia redes neuronais profundas (DNNs) a reconhecer padrões e a reduzir erros ajustando conexões internas"[39]. Este processo requer muitos dados, razão pela qual as celebridades são as mais frequentemente visadas, devido à abundância de material visual disponível para essa operação.

A conceção de *deepfakes* depende ainda de um codificador automático que, comprime imagens detalhadas em características-chave num "espaço latente" (como a pose, expressão ou luz) e depois reconstrói-as. O codificador comprime a imagem em medições essenciais, enquanto o descodificador a recria. Este processo permite gerar rostos realistas e novos a partir de padrões aprendidos.

Resumindo, o segredo dos *deepfakes* é "usar um codificador partilhado para comprimir características comuns de dois rostos diferentes num espaço latente"[40]. Assim, imagens de pessoas distintas, com a mesma expressão ou pose geram medições semelhantes. Isso permite transformar o rosto de uma pessoa no de outra, mantendo a expressão e a postura da imagem original, criando uma imagem falsa, mas realista.

Com base nesta exposição inicial, somos capazes de distinguir este fenómeno de um conceito algo similar, que, no fundo, se interligam: as *fake news* (notícias falsas), que facilmente são confundíveis. Tal como os *deepfakes,* também as *fake news* têm ameaçado a nossa sociedade, em especial os meios de comunicação e a democracia. Histórias falsas são partilhadas e divulgadas com uma velocidade tremenda, particularmente em redes sociais, influenciando inúmeros internautas e as suas crenças e opiniões. "Atualmente, um em cada cinco utilizadores de redes sociais obtém informação a partir do Youtube e do Facebook"[41], o que demonstra o poder destes meios.

Como os vídeos se tornaram bastante populares, é imprescindível verificar se o que estamos a ler ou a ver é autêntico, já que é relativamente fácil manipular vídeos de forma convincente. E, apesar da disseminação de *fake news* ser fácil, é complexo combater os *deepfakes*, por todas características que já vimos.

[38] Ibidem em 34.
[39] Ibidem em 35.
[40] Ibidem em 35.
[41] Westerlund, 2019.

Tal como todas as tecnologias, os *deepfakes* também têm um lado positivo recheado de oportunidades. Por exemplo, partilhando as nossas próprias fotografias e os nossos descodificadores pessoais, é possível criarmos um *deepfake* em que aparecemos num filme de Hollywood, num campo de futebol com o nosso jogador preferido, como se estivéssemos a jogar um jogo de consola, ou um manequim virtual que experimenta várias roupas em nós próprios. É a "derradeira personalização"[42], feita para puro entretenimento e completamente inofensiva. O avanço tecnológico parece trazer à tona tanto o melhor quanto o pior nas pessoas, impulsionando-nos a progredir enquanto também nos faz retroceder simultaneamente no tempo.

Todavia, parece que a maioria dos exemplos de *deepfakes* que encontramos *online* destinam-se à prática de ilícitos, como a violação de direitos de autor e o roubo de identidade, a pornografia infantil, o *revenge porn*, a difamação e, por conseguinte, à lesão de direitos fundamentais e os Códigos Penais, tanto a nível nacional, como internacional, não conseguem lidar com a natureza complexa dos *deepfakes*[43]. Para comprovar esta posição, recorremos ao caso de uma jovem estudante de Direito, uma cidadã desconhecida que, após uma simples pesquisa, encontrou imagens e vídeos dela com a sua cara no corpo de atrizes pornográficas, algo que não só colocou a sua reputação em risco, mas também a sua carreira profissional, bem-estar emocional e segurança. Tal evidencia que qualquer um de nós pode ser vítima do uso de deepfakes[44][45].

---

[42] Dietmar, 2019.

[43] Neste sentido, refira-se o lançamento da Diretiva 2024/1385 do Parlamento Europeu e do Conselho de 14 de maio de 2024, relativa ao combate da violência contra as mulheres e à violência doméstica. No respetivo artigo 5.º, é punível tanto a divulgação ao público como a produção, manipulação ou adulteração de imagens, vídeos ou materiais semelhantes que deem a ideia de uma pessoa estar a participar em atos sexualmente explícitos, sem o seu consentimento. Apesar de a conduta que se pretende criminalizar seja tão somente a participação nos referidos atos e não alguém que simplesmente apareça nua, a União Europeia deu um passo substantivo na criminalização da partilha não consensual de material íntimo ou manipulado por tecnologias de Inteligência Artificial Generativa.

[44] Melville, 2019.

[45] Esta situação configura um de inúmeros casos infelizes que sublinha o potencial negativo dos *deepfakes* em contextos pessoais e profissionais. Também em abril de 2024, Eric Eiswert o diretor da Pikesville High School em Maryland, foi vítima de uma gravação em formato áudio *deepfake.* A gravação foi criada por um antigo diretor desportivo da escola que clonou a voz de Eiswert a proferir comentários racistas e anti-semitas. Como consequência, o mesmo foi suspenso administrativamente, o seu trabalho foi posto em causa e recebeu ameaças de morte. Mais tarde, a gravação foi cuidadosamente analisada por profissionais e o tribunal concluiu que tinha vestígios de conteúdo gerada por Inteligência Artificial, sendo o antigo diretor desportivo da escola detido e acusado de perseguição, roubo e perturbação do funcionamento da escola (Finley, 2024).

Em particular, a Coreia do Sul tem enfrentado uma crise de deefakes, principalmente no que respeita à pornografia não consentida. Estudantes, professores e jornalistas estão a descobrir que estão a ser alvo desta epidemia em rápido crescimento. Os relatórios indicam um aumento de casos, com 297 registados nos primeiros sete meses de 2024, contra 180 em 2023. Em setembro de 2023, um caso específico envolveu um homem condenado por utilizar a Inteligência Artificial para gerar imagens de abuso sexual de crianças, marcando a primeira condenação deste tipo no país. Mais de 500 escolas e universidades estão a ser alvos, com muitas vítimas menores de idade, o que mostra a dimensão do problema e o seu impacto na comunidade (Gong, 2024).

b. Os *Deepfakes* e o Processo Penal: contaminação da verdade probatória

De seguida, iremos debruçar-nos sobre o impacto que os *deepfakes* podem ter efetivamente nas provas reais, no Direito Processual Penal Português e as potenciais estratégias para excluí-los do conteúdo probatório.

A psicologia tem-nos ensinado que os seres humanos valorizam a perceção visual acima de todos os outros indicadores. Historicamente, "o sistema jurídico tem favorecido a admissão de provas audiovisuais. Estudos demonstram igualmente que "os jurados que ouvem testemunhos orais acompanhados de testemunhos em vídeo têm 650% mais probabilidade de reter a informação"[46]. É simples concluir que as provas em vídeo afetam poderosamente a memória humana e a sua perceção da realidade.

Os tribunais vão procurar que estes conteúdos mostrem uma representação fiel e exata da realidade. Assim, naturalmente, podemos considerar uma testemunha não apenas aquela que registou o momento, mas também aquela que o presenciou e é capaz de tecer considerações justas e exatas sobre as circunstâncias daquele material. Para afastar qualquer manipulação, "podemos recorrer a testemunhas que reconheçam o arguido, a sua voz e o lugar representado". Assim, o juiz formará uma convicção objetiva aquando da apreciação da prova. A forma mais simples até de refutar um vídeo *deepfake* acaba por ser a pessoa nele visado testemunhar sob juramento que o respetivo vídeo é falso. Contudo, sem sempre isto é tão linear como se possa pensar.

Todavia, quando uma testemunha não está presente, os juízes só podem confiar em vídeos e imagens como únicas "testemunhas" – "teoria da testemunha silenciosa"[47]. E, como a admissibilidade e a credibilidade começam a ser abaladas, há quem defenda que "as provas baseadas em vídeo e imagem podem começar a ter menos informação e valor do que comportavam anteriormente".[48]

Além da prova testemunhal, para evitar que provas falsas corrompam o processo penal, será fulcral recorrer não só a peritos forenses digitais especializados na área, como a sistemas preparados de deteção de falsificações alimentados pela IA, nos casos em que a qualidade dessa adulteração não seja fácil de identificar, dado os avanços da tecnologia neste fenómeno.

Vale a pena destacar sucintamente algumas contribuições das instituições líderes mundialmente e as e as técnicas mais eficazes que permitem compreender como estes avanços apoiam o sistema de justiça penal. Universidades como a Academia Chinesa de Ciências, a Universidade *Sun Yat-sen* e a Universidade de *Nanjing* estão na linha da frente da pesquisa sobre a identificação de *deepfakes* - essas organizações conceberam algoritmos avançados para reconhecer e reduzir os efeitos de falsificações profundas. Entre as técnicas mais utilizadas estão as redes neurais convolucionais, métodos de

[46] Delfino, 2023.
[47] Breen, 2021.
[48] Fallis, 2020.

aprendizado automático, *Xception*, *3D CNN* e *EfficientNet*. Estas tecnologias focam-se principalmente na deteção desses conteúdos falsos de inconsistências em dados de vídeo, áudio e imagem que possam indicar manipulação[49].

"Será possível erradicar quaisquer suspeitas de manipulação porque serão facilmente descobertas com o desenvolvimento destes sistemas e porque o proponente do vídeo não consegue responder a perguntas básicas para o autenticar (quando e quem criou o vídeo e com que tecnologia", segundo afirma Grant Fredericks, presidente da *Foresinc Video Solutions*.

D'Alessandra e Sutherland escrevem que devemos efetuar uma avaliação do valor probatório dos diferentes elementos de prova, baseando-se na confirmação e na credibilidade das pessoas que os apresentam; além disso, se várias fontes independentes corroborarem essas provas, o vídeo ou imagem em questão passa a ter um peso probatório significativo[50]. Assim, o risco de estramos perante uma prova manipulada decrescia substancialmente.

Não será de subestimar o potencial lesivo dos *deepfakes*. Em 2010, um tribunal de recurso do Estado da Califórnia recusou a autenticidade de uma foto descarregada de uma rede social pelo facto de não ter sido ratificada por um depoimento de uma testemunha ou por um parecer de um perito. O tribunal sublinhou a importância da existência deste mesmo parecer uma vez que "as fotografias digitais podem ser alteradas para produzir imagens falsas e co advento de programas informáticos como o *Adobe Photoshop*, nem sempre é necessária experiência ou mesmo conhecimento para alterar uma fotografia digital"[51]

Nos EUA, os processos judiciais da última década mostram que, apesar dos avanços na tecnologia *deepfake* e no risco de manipulação de áudio e vídeo, os tribunais mantêm regras liberais de autenticação de provas digitais e rejeitam critérios mais rigorosos. Casos recentes (2014-2019) reforçam esta abordagem, rejeitando a presunção de que todos os vídeos digitais sejam suspeitos por padrão. Os mecanismos desenvolvidos ao longo dos anos devem ser eficazes contra *deepfakes*, assim como foram contra tecnologias anteriores (lembrando que a manipulação de imagens é um risco conhecido desde os primeiros anos da fotografia).

No que ao Direito Processual Penal Português diz respeito, é indiscutível que os *deepfakes* têm o potencial de comprometer o conjunto de provas e influenciar a decisão final do juiz. Desde a fase inicial do processo, em que se realizam investigações para determinar se ocorreu um crime e quem são os seus responsáveis, passando pela decisão sobre medidas de coação, como a prisão preventiva, até à conclusão do julgamento e à sentença, a prova desempenha um papel fundamental para garantir uma decisão justa e adequada, conforme o artigo 124.º do Código de Processo Penal (CPP). De acordo com Germano Marques da Silva, "a função essencial do processo penal é decidir se um crime foi cometido,

[49] Sandoval, de Almeida Vau, Solaas e Rodrigues, 2024.
[50] D'Alessandra & Sutherland, 2021.
[51] Acórdão de 24 de junho de 2010, Tribunal de Recurso da Califórnia, Segundo Distrito. 185 Cal.App.4th 509 (Cal. Ct. App. 2010).

determinar os seus autores e apurar a responsabilidade criminal. Para atingir esse objetivo, a prova desempenha um papel crucial, pois, segundo Bentham, o processo penal não é mais do que a arte de administrar a prova"[52].

Neste sentido, nos termos do artigo 127.º do Código de Processo Penal, surge o princípio da livre apreciação da prova, que estipula que "a prova é apreciada segundo as regras da experiência e a livre convicção da entidade competente".

Assim, este princípio tem uma dupla vertente: dignifica, negativamente, a falta de regras legais que estabeleçam previamente o valor da prova e, de forma positiva, que as autoridades responsáveis pela avaliação da prova o façam com o compromisso de buscar a justiça e a verdade material, realizando uma análise que deve ser sempre objetiva, fundamentada e, portanto, passível de revisão.[53]. O que indica que o juiz de julgamento, o juiz de instrução e o Ministério Público devem então valorar a prova não de forma predeterminada pela lei, mas de acordo com as regras de experiência e a sua livre convicção[54].

Não se trata, contudo, de um princípio sem exceções, como o demonstram, por exemplo, as regras relativas ao depoimento indireto (artigo 129.º do CPP) ou às vozes públicas e convicções pessoais (artigo 130.º do CPP), mas aplica-se sem restrições no que toca à apreciação de reproduções fotográficas, cinematográficas ou fonográficas.

O tribunal deve ordenar "oficiosamente ou a requerimento, a produção de todos os meios de prova cujo conhecimento se lhe afigure necessário à descoberta da verdade e à boa decisão da causa", conforme o artigo 340.º, número 1 do Código de Processo Penal (princípio da investigação). Desta forma, o tribunal examina o facto em julgamento por conta própria, sem depender apenas dos argumentos da acusação ou da defesa, formando de maneira independente as convicções e o fundamento para a sua decisão.

Com base nestas normas, neste capítulo propomo-nos a compreender como este fenómeno dos *deepfakes* poderá afetar o processo penal e se tornaram obsoletos os atuais princípios elementares do processo penal português e, ao mesmo tempo, do direito positivo.

A ordem normativa vigente e os tribunais enfrentam um problema complexo e que não deve ser menosprezado: o desconhecimento, muitas vezes generalizado, destas mudanças e realidades atuais. Os *deepfakes* são capazes de "minar a confiança nas instituições (jurídicas e policiais) e o conceito de verdade, dificultar a atribuição e a construção de um caso jurídico e, falsificar o conteúdo probatório em tribunal"[55]. Este último desafio merecerá largamente a nossa atenção, já que é o mais amplamente discutido entre autores e é a ameaça mais significativa.

[52] Carrão, 2022.
[53] Antunes, 2018.
[54] Ibidem em 34.
[55] Ibidem em 49.

Desta forma, a pergunta que se coloca é: como é que estes princípios podem ser aplicados de forma concreta e objetiva se quem avalia as provas se baseia em premissas ultrapassadas, confundindo passado, presente e futuro, e ignorando estes riscos contemporâneos para a sociedade e para a administração da justiça?

A verdade é que os modelos atuais de justiça penal, mesmo quando funcionam de maneira ideal, não conseguem assegurar completamente o uso amplo de provas altamente confiáveis e avançadas. Cada novo tipo de prova levanta dúvidas sobre como deve ser corretamente incorporado nos processos penais, e à medida que os tribunais passam a lidar com provas cada vez mais complexas e inovadoras, surgem novos desafios que devem serem superados. "Inexistindo uma robusta literacia digital que acompanhe o domínio da criação, interpretação e aplicação do direito, abrem-se as portas para um fenómeno de obsolescimento material do direito"[56].

Juridicamente, os *deepfakes* são uma temática ainda embrionária. As mais recentes investigações prendem-se com a questão da contenção, ou seja, como evitar e responder ao uso indevido dessa tecnologia para fins nocivos e de que forma a legislação deve reagir quando os *deepfakes* provocam prejuízos, seja a nível individual ou em grande escala, como também na segurança nacional[57].

Como adverte a Europol, as Nações Unidas e a Trend Micro, o conteúdo audiovisual *deepfake* pode ser apresentado de forma maliciosa como prova legítima na tentativa de frustrar investigações criminais e processos judiciais, lançando dúvidas sobre as provas audiovisuais como uma categoria de provas[58]. E os efeitos serão diretos e indiretos. Por um lado, este fenómeno dos *deepfakes* aumentará o número de processos nos tribunais, com a possibilidade de poderem dar origem a responsabilidade civil. Por outro lado, se os *deepfakes* poderão servir como uma prova processual, tal pode levantar vários constrangimentos e afetar a credibilidade dos processos judiciais e a respetiva audiência: os advogados que tentarão introduzir ou excluir vídeos como prova, os juízes que decidem e indicam se o vídeo é ou não admissível e testemunhas e os peritos que serão chamados a depor sobre o respetivo vídeo.

Exige-se uma resposta multidisciplinar a estes desafios apresentados pelos *deepfakes*, "que converge na aquisição de uma elevada literacia digital das pessoas com um enquadramento jurídico adequado"[59]. Um devido investimento na educação garante que o acesso às novas tecnologias seja complementado com uma abordagem crítica à quantidade de informação que vamos tendo acesso. Esta capacidade de lidar com os *deepfakes* de forma responsável deve partir não só da sociedade, mas também das empresas e das novas ferramentas criadas, com vista detetá-los eficazmente.

[56] Ibidem em 34.
[57] Pfefferkorn, 2020.
[58] Ibidem em 34.
[59] Floridi, 2018.

i. A fragilidade dos inimputáveis

Não deveremos também esquecer que a introdução de provas geradas por Inteligência Artificial nos tribunais criminais tem suscitado debates sobre a equidade processual, no que diz respeito aos arguidos inimputáveis, que já vêm sendo referidos ao longo do trabalho - alguém que, por força de uma anomalia psíquica, for incapaz, no momento da prática do facto, de avaliar a ilicitude deste ou de se determinar de acordo com essa avaliação e quem, por força de uma anomalia psíquica grave, não acidental e cujos efeitos não domina, sem que por isso possa ser censurado, tiver, no momento da prática do facto, a capacidade para avaliar a ilicitude deste ou para se determinar de acordo com essa avaliação sensivelmente diminuída, segundo o artigo 20.º, números 1 e 2, respetivamente.

Devido à sua condição, estes indivíduos podem ter dificuldade em compreender os procedimentos legais ou comunicar eficazmente em tribunal e/ou com o seu advogado. Isto pode dificultar a sua defesa, especialmente quando as provas são alvo de manipulação, como falsos testemunhos ou documentos alterados, ou mesmo em depoimentos de testemunhas pouco fiáveis. Esta vulnerabilidade quando confrontados com *deepfakes* pode dar lugar a resultados injustos, sendo que a comunicação em tribunal se pretende que seja clara.

Assim, estes arguidos enfrentam barreiras significativas para contestar provas baseadas em IA Generativa devido à complexidade técnica, falta de transparência e limitações inerentes à sua condição.

c. A Prova Pericial

Podemos contar que um certo ceticismo comece a abraçar o processo judicial na medida em que manipulações profundas de vídeos, áudios ou fotos surjam no contexto probatório. Uma parte pode produzir um vídeo para efeitos de litígio ou mesmo encontrar um *deepfake* e querer apresentá-lo como prova sem se aperceber do que verdadeiramente do que se trata. Mesmo em situações que não envolvam vídeos falsos, a mera existência de *deepfakes* causará incerteza na tarefa de declarar autênticas provas reais. Alguém que argumente a não fidedignidade de um vídeo que não foi objeto de manipulação semeará a dúvida no juiz, no mínimo. É de esperar que as partes recorram a peritos ou a plataformas apuradas para provar cabalmente a veracidade daquela prova.

Mais cedo ou mais tarde, a verdade é que os tribunais sempre se mostraram instituições capazes de lidar com o velho problema da falsificação e com as várias reinvenções do Direito e a sua resiliência deve atenuar alguma histeria ao redor dos *deepfakes*.

Uma forma de impedir a instrumentalização do sistema de justiça penal por *deepfakes* seria estabelecer regras de admissibilidade para provas fotográficas, cinematográficas ou fonográficas, atribuindo ao sujeito que as apresenta, especialmente o assistente ou arguido, o ónus de comprovar sua autenticidade, como veremos *infra*. Isso exigiria a apresentação de um relatório pericial confirmando, com base na tecnologia atual, a ausência de indícios de manipulação. Para evitar críticas de parcialidade, a

responsabilidade por esses relatórios poderia ser atribuída a instituições ou serviços oficiais, como já foi previsto para outras provas periciais nos termos do artigo 152.º do CPP.

Embora a lei não defina explicitamente o que é a perícia, limitando-se a indicar as situações em que a mesma deve ser realizada, Germano Marques da Silva defini-a como "atividade de perceção ou apreciação dos factos probandos efetuada por pessoas dotadas de especiais conhecimentos técnicos, científicos ou artísticos"[60].

Debruçando-nos sobre o regime legal, a perícia está regulada no nosso Código de Processo Penal nos artigos 151.º a 163.º e tem lugar quando a perceção ou apreciação dos factos exigir especiais conhecimentos técnicos, científicos ou artísticos (artigo 151.º). É realizada em estabelecimento, laboratório ou serviço oficial apropriado (artigo 152.º, número 1). O artigo 153.º refere-se à função de perito e, nos termos do disposto do artigo 154.º, número 1, a perícia é ordenada, oficiosamente ou a requerimento, por despacho da autoridade judiciária, contendo a indicação do objeto da perícia e os quesitos a que os peritos devem responder. A competência para ordenar a perícia é atribuída ao Ministério Público durante a fase de inquérito, ao Juiz de Instrução na fase de instrução e ao juiz durante a fase de julgamento.

A prova pericial traria a vantagem de aumentar a confiança e a tranquilidade na valoração de provas, mas também poderia gerar certas desvantagens tais como a possibilidade de essa mesma confiança se ter alicerçado num falso negativo ou num falso positivo, ou seja, quando os *deepfakes* não são detetados, devido ao seu elevado grau de realismo ou utilizam técnicas avançadas para evitar a deteção ou quando os algoritmos de deteção ou os peritos identificam erradamente um conteúdo genuíno como um *deepfake*, e devido a artefactos nos meios de comunicação ou a limitações no modelo de deteção, respetivamente.

Isto salientaria o potencial lesivo dos *deepfakes* ao ter passado pelo crivo da perícia e o comprometimento da integridade dos julgamentos penais, afetando os resultados e conduzindo potencialmente a erros na decisão.

Além disso, surge ainda a questão de quem arcaria com os custos da mesma, no sentido de que, se recaísse sobre quem apresentasse a prova, poderia desincentivar o seu uso. O resultado teria um impacto desproporcionado no acesso à justiça, sendo que as partes socioeconómicas mais vulneráveis terão esse caminho mais dificultado, saindo favorecidas aquelas mais abastadas. Ou seja, a falta de acesso equitativo à perícia forense pode significar que a capacidade de um indivíduo para provar a sua inocência ou contestar evidências pode depender de sua condição socioeconómica, reforçando desigualdades já existentes no sistema legal.

É sabido que a autenticação de *deepfakes* exige métodos forenses avançados, como a análise de integridade digital, deteção de inconsistências em padrões de iluminação e rastreamento de metadados e todas essas análises dependem de softwares sofisticados e

[60] Silva, 2008.

de especialistas altamente treinados, cujos custos que lhes estão inerentes podem facilmente constituir um impeditivo para pessoas com baixo poder financeiro. Por outro lado, indivíduos mais ricos ou corporações têm capacidade pagar por peritos conceituados para autenticar ou contestar provas digitais.

A disparidade no acesso a tecnologias de deteção de *deepfakes* pode assim resultar numa "justiça de dois pesos e duas medidas", em que acusados com poucos recursos podem mesmo vir até a ser condenados com base em provas falsificadas, em *última ratio*, sem possibilidade de as questionar.

Conforme escreve o *Financial Times*, o advento de *deepfakes* também deu origem ao chamado "liar's dividend", (o dividendo do mentiroso), em que "indivíduos que são apanhados em imagens genuínas e comprometedoras afirmam que são falsas, semeando assim a dúvida e evitando a responsabilização"[61]. Esta tática pode ser particularmente vantajosa para quem tem recursos para corromper peritos, reforçando ainda mais as disparidades entre as partes ricas e mais desfavorecidas em contextos processuais.

Do mesmo modo, a Inteligência Artificial pode "incentivar políticos a mentir sobre a autenticidade de conteúdos reais"[62]. A crescente consciencialização de pública sobre *deepfakes* pode ser perfeitamente explorada por figuras públicas para desacreditar os cidadãos sobre certos conteúdos, que, na realidade, são verdadeiros. Este comportamento tem implicações nefastas em casos eleitorais, por exemplo, minando a integridade eleitoral e a confiança pública.

Este problema foi explorado por uma autora norte-americana num estudo na qual a mesma refere que em casos que envolvam *deepfakes* ou provas geradas por Inteligência Artificial, a falta de acesso a perícias digitais especializadas devido ao elevado custo que acarretam pode afetar desproporcionalmente minorias e pessoas economicamente mais vulneráveis. Sem esta capacidade financeira, estas pessoas podem ser privadas de defender os seus direitos adequadamente.

Exemplo disto foi o caso de uma mãe chamada Raffaela Spone, conhecida como "*deepfake cheerleader mom*"[63], que "foi presa por várias acusações de assédio por, alegadamente, ter criado imagens deepfakes das líderes da claque adversária da sua filha com o propósito de as atacar, publicando vídeos das mesmas a beber, fumar, fazendo-as também parecer nuas".

Este caso tornou-se mediático e Spone recebeu imensas ameaças de morte, enquanto negava afincadamente ter criado aqueles *deepfakes.* De acordo com a sua advogada, ela não tinha recursos financeiros para pagar pela devida perícia de modo a provar a sua inocência. Porém, uma empresa especializada em tecnologia voluntariou-se para prestar os seus serviços de forma pro bono e determinou a inautenticidade dos vídeos anteriormente criados.

[61] Murphy, 2024.
[62] Goldenstein e Lohn, 2024.
[63] Delfino, 2024.

A lei não prevê qualquer salvaguarda para este tipo de situações. Cada parte suportará as suas despesas no que à apresentação de provas diz respeito, bem como os honorários dos advogados.

i. Olhar sobre os Modelos de Perícia

Nesta matéria, conforme a ordem jurídica, podem ser considerados dois modelos de perícia digital: o Sistema de Perícia Oficial do Estado, ou o Modelo de Perícia Independente, em que a mesma é contratada por advogados.

No primeiro modelo, o Estado arca com os custos da perícia, garantindo que todas as partes têm acesso a especialistas qualificados, independentemente de sua condição socioeconómica. Este sistema existe em diversos países, como o Brasil e a Alemanha e caracteriza-se pela imparcialidade e neutralidade dado que os peritos designados atuam como auxiliares da justiça, não estando vinculados a nenhuma das partes envolvidas no processo[64]. O Estado supervisiona e controla os resultados das perícias, garantindo que os peritos seguem as normas estabelecidas.

Uma das características mais importantes deste modelo é que a perícia oficial é gratuita para as partes envolvidas no processo judicial. Portanto, o Estado arca com os custos das perícias, garantindo que todos, independentemente da sua capacidade financeira, têm acesso aos serviços da perícia.

Quanto ao modelo de perito privado, característico do ordenamento jurídico norte-americano, "hoje, cada parte está encarregue das suas próprias despesas de litígio, a menos que uma lei, um contrato ou uma regra transfira esse ónus para a parte vencida"[65]. Claro está que apesar de ser contratado pelas partes, ao perito privado cumpre atuar com independência, sem se influenciar pelos interesses da parte que o contratou. Além disso, nenhum dos mecanismos de atribuição de custos cobre adequadamente os custos de peritos associados a provas falsas para aqueles que não têm meios para pagar esses custos.

Já o sistema português é designado como "misto ou contraditório mitigado, por não serem as partes quem designam os peritos, havendo prevalência de intervenção de organismos públicos, ainda que seja permitido às partes, subsidiariamente, em caso de inexistência desses, a apresentação de perícias contraditórias"[66][67]

Tal como foi apresentado, provar ou refutar *deepfakes* exigirá provavelmente provas periciais dispendiosas e alguns litigantes considerarão estes custos impossíveis de suportar. Manter o acesso à justiça em casos que envolvam falsificações profundas exige

[64] Zambon Perícia, 2021.
[65] Ibidem em 63.
[66] Veríssimo, 2023.
[67] O Estado vai suportar o pagamento da perícia, que é antecipado pelo tribunal. Contudo, esse pagamento vai entrar em regra de custas processuais, sendo pagas afinal caso haja lugar ao respetivo pagamento de custas. Nos termos dos art. 16.º, número 1, alínea a) e 17.º do Regulamento das Custas Processuais, aprovado pelo Decreto-Lei n.º 34/2008, de 26 de fevereiro.

uma "resposta multidimensional que obriga à alteração dos estatutos relativos aos custos e regras processuais"[68].

Uma das soluções apresentadas por Rebecca Delfino (para as ordens jurídicas em que prevalece o sistema de perito privado) passa pelo "alargamento das categorias de custos recuperáveis para incluir todos os honorários e custos de peritos, através de uma alteração da lei", dada o incremento considerável de provas audiovisuais que são apresentadas em processos judiciais e a sua respetiva dificuldade de análise.

Outra possibilidade consistiria em ser "presumivelmente atribuído à parte que levanta a alegação de *deepfake,* quer ofensivamente quando procuram impor a responsabilidade à outra parte, como defensivamente como esforço para se defenderem"[69]. Isto porque alegar que se trata de um *deepfake* é simples de afirmar, mas oneroso para provar. Assim será será necessário cautela e ponderação antes alegar de que algum conteúdo é manipulado. Porém, esta presunção pode ser afastada se a parte que levanta a questão do *deepfake* comprovar a sua falta de recursos financeiros para arcar com os custos da peritagem.

Já Luisa Verdovila sugere a necessidade de criar sistemas ou bases de dados de *deepfakes* acessíveis, que possam ser usados tanto por profissionais do direito quanto pelo público em geral para comprovar a autenticidade de provas digitais. Estes recursos devem estar junto das instituições judiciais, garantindo que a verificação técnica não depende somente de especialistas mais dispendiosos[70].

Desta forma, resta-nos concluir que a resposta mais imediata aos *deepfakes* passaria pela produção de prova pericial, conforme os artigos 151.º e seguintes do CPP, devido à complexidade técnica de identificar conteúdos sintéticos, que exigem conhecimentos que estão numa esfera que o jurista comum naturalmente não predispõe.

O juízo sobre a necessidade de prova pericial não é plenamente livre e esclarecido se o julgador desconhece os fundamentos tecnológicos que sustentam o pedido. E delegar a solução a um especialista exige, antes de tudo, reconhecer a complexidade do problema.

É verossímil que os *deepfakes* irão ameaçar o processo penal e fazer questionar a atualidade dos princípios basilares do direito processual penal do nosso ordenamento jurídico. Urge alguma diligência na verificação da autenticidade das provas digitais de modo a "separar o joio do trigo", reduzindo ambiguidades de modo que o juiz possa apreciá-las e decidir livremente e objetivamente.

Assim, exige-se ainda uma abordagem cuidadosa e multidisciplinar para garantir a integridade do sistema de justiça penal e evitar a sua instrumentalização. A imposição de regras de admissibilidade para provas audiovisuais e a exigência de uma perícia técnica para comprovar a sua autenticidade representam um avanço essencial, mas também

[68] Ibidem em 63.
[69] Ibidem em 63.
[70] Verdovila, 2020.

levantam questões e desafios significativos relacionados com a equidade no acesso à justiça.

A disparidade financeira no acesso à perícia forense digital pode gerar uma justiça desigual, na qual aqueles com maiores recursos financeiros conseguem refutar ou corroborar provas de maneira mais eficiente, enquanto os menos abastados enfrentam dificuldades para garantir a sua defesa. Esta assimetria reforça desigualdades já existentes no sistema jurídico e pode resultar em condenações erróneas baseadas em provas manipuladas, sem possibilidade de uma defesa adequada e digna.

Portanto, a regulamentação e a democratização da perícia digital são fundamentais para impedir que os *deepfakes* distorçam a justiça penal. A criação de mecanismos transparentes e acessíveis para a verificação de provas digitais deve ser uma prioridade para garantir que todos, independentemente de sua condição socioeconómica, tenham igualdade no acesso a um julgamento justo e imparcial.

4. A integração da Inteligência Artificial Generativa no quotidiano e soluções para este fenómeno, em especial no mundo jurídico

Como vimos, a IA Generativa é idealizada para a criação de conteúdo escrito e fotográfico em poucos minutos ou até segundos, o qual, com o esforço humano, demoraria dias ou até semanas, aumentando substancialmente a produtividade em diversos setores, otimizando os seus processos. Com esta capacidade que supera os processos cognitivos humanos, consegue gerar textos com rapidez e precisão, permitindo que as empresas poupem tempo e recursos.

Mas no que ao Direito diz respeito, a Inteligência Artificial não foge à regra e também pode ser utilizada para diferentes fins, tais como "a produção de textos para contratos, testamentos, trabalhos académicos ou simplesmente para acelerar uma simples pesquisa jurídica"[71].

Uma área com particular potencial de impacto da IA Generativa prende-se com a redação de requerimentos ou preparação de peças processuais. Ao litigar um caso, um profissional de justiça pode apresentar um pedido para arquivar ou suspender um caso, para pedir que uma parte intervenha, para proibir um determinado testemunho, entre outras. Não há dúvidas de que o prazo de elaboração deste pedido ou de qualquer outra peça pode ser curto, exigindo ainda uma certa completude.

É errado confiar nos atuais sistemas de IA Generativa para gerar peças e textos prontos à sua submissão. Isto porque um dos desafios que estes modelos enfrentam são o risco de "alucinações", em que gera informação plausível e estruturada, mas factualmente incorreta ou sem sentido. Têm sido realizados vários esforços na tentativa de redução e mesmo e respetiva eliminação deste inconveniente característico destes modelos. Porém, há estudos que afirmam que é impossível extinguir completamente as "alucinações" dos

[71] Ibidem em 2.

LLM. Um desses estudos mostra-nos que, definindo a alucinação como uma discrepância entre as respostas de uma LLM computável e uma função de verdade computável, "recorrendo a resultados da teoria da aprendizagem, os LLM não podem aprender todas as funções computáveis e, por conseguinte, terão sempre alucinações"[72].

No entanto, é razoável fornecer e garantir ao sistema de IA o acesso a um conjunto completo de documentos e jurisprudência, tais como decisões ou recursos sobre temáticas semelhantes ao processo em questão, que estejam disponíveis publicamente, para que esse mesmo sistema de IA crie um rascunho inicial, em vez dessa versão final, podendo contribuir decisivamente e intelectualmente, resultando num esforço conjunto com o advogado.

Desta forma, quase de imediato, esse profissional receberia um esboço produzido à luz das informações referidas. Embora esse esboço não seja perfeito e necessite de ser verificado a nível de direito e de facto e reformulado, estará disponível prontamente. É provável que num espaço temporal de cinco, dez ou quinze anos, com os avanços tecnológicos, as empresas criem sistemas de IA para redigir documentos jurídicos que incluam uma funcionalidade que verifica automaticamente todas as citações e afirmações presentes no documento com o objetivo de garantir a sua exatidão. Com este processo, o profissional de justiça "pode ser capaz de produzir numa tarde um documento que teria exigido vários dias ou semanas de redação e de revisão sem a IA"[73].

Não obstante o supramencionado, mesmo com anos de avanços na IA Generativa, um bom advogado continuará a ter uma visão mais completa e profunda de um caso do que qualquer sistema de inteligência artificial. É fundamental não permitir que a ferramenta ofusque o verdadeiro objetivo da tarefa. A IA generativa pode ser extremamente poderosa, mas, no final das contas, não deixa de ser apenas um instrumento de apoio[74].

Há muitas empresas que se encontram a trabalhar no desenvolvimento de produtos de IA Generativa. Em Portugal, um desses produtos foi lançado no mês de outubro de 2024 – a *LeiPT*, que se autointitula como uma "assistente pessoal com conhecimento de legislação e jurisprudência portuguesa", [como] "toda a legislação consolidada do Diário da República, acórdãos do Supremo Tribunal, Tribunais da Relação, Tribunais Administrativos, Tribunal dos Conflitos, pareceres da PGR, CAAD", entre outras. [Através desta plataforma, é possível] "discutir argumentos e estratégias para de casos, formular acordos extrajudiciais de forma rápida e eficaz, encontrar artigos e acórdãos relevantes ou automatizar o processo de redação de documentos"[75], tudo em meros segundos.

Além disso, no âmbito da *Web Summit*, uma das maiores cimeiras tecnológicas internacionais, em novembro de 2024, o nosso atual Primeiro-ministro anunciou na sua abertura o lançamento de um modelo de linguagem de IA em português, considerando

[72] Xu & Jain & Kankanhalli, 2024.
[73] Ibidem em 2.
[74] Ibidem em 2.
[75] LeiPT, 2024. Disponível em https://www.lei-ai.pt/

este “um passo crítico" para o ensino, administração pública ou empresas”[76] – o *LLM Amália* -, tendo como propósito a promoção de tutores de IA para a educação adaptada, serviços públicos simplificados e personalizados, e o apoio a empresas na transição para a era da IA em português. Luís Montenegro afirmou que este modelo estaria pronto no primeiro trimestre de 2025. Porém, tratar-se-á de uma versão incompleta, com o objetivo de recolha de feedback para o seu aprimoramento, sendo a versão final lançada em 2026.

Posto isto, e entrando numa vertente académica e profissional, uma das questões que tem causado mais controvérsia é o facto de os escritórios de advogados e as faculdades de Direito e os seus alunos poderem envolver-se na conveniência de utilizar o *ChatGPT* ou outra plataforma de IA Generativa e se a proibição constitui a solução mais acertada a este desafio.

A discussão acentuou-se aquando de uma das últimas versões do *ChatGPT*, o *ChatGPT-4,* ter passado o Exame da Ordem dos Advogados Norte-Americano. O que mais surpreendeu foi que o *ChatGPT-4* não só se limitou a passar, como obteve bastante sucesso na secção de escolha múltipla do exame e também em ambas as partes da componente escrita, superando a média dos candidatos reais e alcançando uma pontuação de 90%.[77]

No que aos escritórios de advogados diz respeito, o impacto da IA Generativa irá depender numa multiplicidade de fatores. Há um “interesse inequívoco em seguir os desenvolvimentos tecnológicos inerentes ao Direito e em providenciar os seus profissionais as melhores ferramentas para que eles possam trabalhar de forma mais eficiente”[78].

Por um lado, é fulcral ter em mente que o fluxo de mercado a nível tecnológico está constantemente em mutação, o que significa que rapidamente as primeiras plataformas de IA Generativa poderão tornar-se-ão obsoletas ou mesmo desaparecer. Além disso, também se vão revelar dispendiosas em relação ao que estará disponível mais tarde no mercado jurídico, em termos de euros e em termos de custos de formação de literacia digital.

Por outro lado, consideramos um erro um escritório excluir completamente a IA dado que, como já foi diversas vezes referido, oferecerá vantagens que dificilmente serão passíveis de desmerecer a nossa atenção com o amadurecimento da tecnologia, a longo prazo. Quem incorrer na conduta de não utilizar a IA Generativa terá mais custos em comparação com a concorrência. Assim como hoje seria inconcebível que um escritório de advogados não disponibilizasse ferramentas para pesquisa de jurisprudência, também será impensável, num futuro próximo, que não ofereça aos seus advogados soluções avançadas de IA Generativa[79].

[76] Santos, 2024.
[77] Arredondo, 2023.
[78] Ibidem em 2.
[79] Ibidem em 2.

Será importante observar e estar a par das dinâmicas de mercado e colocar numa balança custos financeiros, considerações de natureza ética, por exemplo, a proteção de dados do cliente e testar aquelas plataformas que se poderão revelar promissoras num longo espaço de tempo.

No que toca às Faculdades de Direito, é inevitável atestar que qualquer recém-licenciado que comece a exercer ativamente alguma profissão jurídica (e não só) terá de conviver com poderosas ferramentas de IA Generativa, as mesmas já fazem parte do nosso universo. Assim, é recomendável que as instituições de ensino preparem os seus discentes para trabalharem com elas e se mostrarem como candidatos mais competentes a qualquer função.

De modo a auxiliar a que os alunos desenvolvam esta *skill,* "as faculdades de direito devem melhorar o seu sistema de pesquisa e de escrita para ensinar especificamente os seus alunos a usar a IA Generativa de forma responsável"[80]. Naturalmente que o termo "responsável" levanta alguma complexidade, daí ser necessário impor algumas regras quanto ao seu uso dado que pode ter um impacto na aprendizagem.

Num estudo efetuado por Jonathan H. Choi e Daniel Schwarcz, dois professores de Direito da Universidade da Califórnia do Sul e de Minnesota, respetivamente, concluíram que o *ChatGPT*-4 "melhorou o desempenho médio dos alunos apenas em perguntas simples de escolha múltipla, sem nenhuma alteração no desempenho em perguntas de escrita". Além disso, "o efeito da utilização da IA diferiu de forma significativa consoante o nível de desempenho inicial dos alunos: aqueles com resultados mais baixos beneficiaram dela consideravelmente, enquanto os de melhor desempenho poderão até ter sido prejudicados pelo uso da IA. Estas conclusões têm implicações importantes para o futuro do Direito e da educação jurídica.[81]

Ainda foi possível verificar que o *ChatGPT-4* sozinho teve um desempenho superior ao dos humanos e ao dos humanos assistidos por IA. A médio prazo, tal levanta a possibilidade de os trabalhadores serem totalmente substituídos em certas tarefas. O facto de esta plataforma ter a capacidade de superar os humanos com acesso à mesma tem implicações nocivas à permanência no mercado de profissionais com funções de análise simples sob o *status quo,* como é o caso de assistentes de escritórios de advogados.

Por outro lado, o facto de o *ChatGPT-4* ter melhorado a *performance* dos alunos com pior desempenho pode abrir portas para uma certa igualdade na prática da advocacia e de qualquer outra profissão, implicações estas que se revelam positivas para a sociedade a nível de oportunidades de carreira, por exemplo.

Finalmente, cumpre explicar a surpreendente questão de os melhores alunos terem tido uma pior prestação com o auxílio da IA. Apesar de haver poucas evidências em como o uso da IA pode afetar a qualidade do nosso trabalho, tudo nos leva a crer que a criatividade é suscetível de sair minada e os alunos facilmente se podem contentar com as repostas

[80] Ibidem em 2.
[81] Choi & Schwarcz, 2023.

rápidas da IA em vez de se esforçarem nas questões que exigem um empenho maior. Há outras investigações que mostram que "o acesso a uma IA de alta qualidade induziu os trabalhadores a esforçarem-se menos, quase como se estivessem num estado de "adormecer ao volante ou ao teclado"[82]. Coloca-se o problema de indagar qual será o grau *otimal* de dependência que qualquer trabalhador deve ter com a IA. Assim como não devemos menosprezar a IA e tomá-la como imprecisa, também não nos parece assertivo deixar que a IA conduza inteiramente o nosso trabalho autonomamente.

Porém, o declínio da *performance* dos melhores alunos também pode ser explicado pela simples razão de que, no mundo hodierno, poucos são os estudantes que têm uma base de conhecimento sólida sobre como usar a IA, ao qual se acresce que estas mesmas plataformas estão em constante desenvolvimento, sendo rudimentares comparativamente àquelas que existirão decorridos alguns anos. Com melhores ferramentas de IA aliadas a uma maior competência dos estudantes na sua utilização, acreditamos que o seu acesso permite melhorar a qualidade do seu trabalho escrito.

As instituições de ensino devem encarar a IA Generativa como uma oportunidade e não como uma ameaça, não incorrendo numa política de proibição dado que a literacia digital se vem tornando cada vez mais essencial por todas as questões que temos visto.

De uma perspetiva puramente tecnológica e prática, proibir a utilização de IA para escrever trabalhos seria problemático no sentido em que, como atualmente não existe um mecanismo fiável para identificar texto gerado pela IA, as escolas que impusessem essa restrição iriam acusar inevitavelmente alguns estudantes que a utilizassem, mesmo que tal realmente não tivesse acontecido.

Ao invés, é aconselhável proporcionarem uma devida formação sobre a utilização apropriada. Os alunos devem ser informados de que são plenamente responsáveis por todo o trabalho escrito redigido por eles, o que inclui a verificação de referências e a confirmação da veracidade das informações apresentadas. [Assim, as instituições de ensino devem supervisionar] e "ajudar os estudantes a utilizar a IA Generativa com plena consciência do contexto, ou seja, para a escrita académica a originalidade é fundamental, ao passo que, para outro tipo de documentos jurídicos, como contratos, acaba por ser menos relevante o critério da unicidade"[83]. Não devemos esquecer é que qualquer aluno ou profissional deve garantir que os documentos são rigorosamente escritos e corretamente citados. O desleixo e o descuido não podem entrar nesta equação.

5. Direito Comparado

A regulação europeia sobre *deepfakes* e Inteligência Artificial é atualmente regida pelo Ato de Inteligência Artificial da União Europeia (*EU AI Act*)[84], publicado no Diário

[82] Sako, 2024.
[83] Ibidem em 2.
[84] Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho (13 de junho de 2024). Disponível em https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689.

Oficial da UE em 12 de julho de 2024, constituindo tal um cenário de evolução com destaque para a proteção e inovação.

O *EU AI Act* impõe obrigações de transparência para sistemas de IA que interajam com pessoas ou geram/manipulam conteúdos. Especificamente, o artigo 50.º, número 4, exige que esses sistemas que geram ou manipulam conteúdos, como os *deepfakes*, informem os utilizadores de que estão interagindo com um sistema de IA (a menos que seja óbvio pelo contexto). Além disso, os prestadores de sistemas de IA que geram conteúdos sintéticos difíceis de distinguir de conteúdos humanos devem incorporar soluções técnicas, como marcas d'água, metadados ou métodos criptográficos, para marcar e detetar o conteúdo em formato legível por máquina, conforme o previsto nos pontos 133 e 134 do Preâmbulo.

De referir ainda o artigo 5.º que proíbe práticas de IA que representem risco inaceitável, incluindo sistemas que usem técnicas subliminares ou enganosas para distorcer o comportamento humano, causando danos significativos (físicos, psicológicos ou financeiros. Isso pode incluir *deepfakes* usados para manipulação comportamental, como influenciar eleições ou extorsão. Além disso, é proibida a criação ou expansão de bancos de dados de reconhecimento facial através da recolha aleatória de imagens faciais a partir da *Internet* ou de imagens de televisão em circuito fechado.

O não cumprimento das respetivas obrigações de transparência presentes no artigo 50.º pode resultar em multas de até 15.000.000 € ou 3% do faturamento anual global do ano financeiro anterior, o que for maior, conforme o artigo 99.º, número 4, alínea g.

No nosso país, a regulação principal para *deepfakes* e Inteligência Artificial é o *EU AI Act*, que impõe transparência para conteúdos gerados por IA, como *deepfakes*, com etiquetagem obrigatória para informar os utilizadores, conforme supramencionado. Todavia, Portugal não possui uma lei nacional abrangente além do *EU AI Act*. Cumpre ainda assim referir a estratégia nacional "AI Portugal 2030", lançada em 2019, que visa "promover e mobilizar a sociedade em geral, para o ensino e investigação, para a inovação e desenvolvimento de produtos e serviços suportadas em tecnologias IA".[85] Esta estratégia está alinhada com o Plano de Ação da UE para IA, mas não inclui disposições legais específicas.

Além do *EU AI Act,* parece não existir uma definição uniforme de IA em Portugal. Porém, as leis portuguesas gerais podem perfeitamente ser aplicadas. O Regulamento Geral de Proteção de Dados (RGPD), que protege dados pessoais, incluindo imagens e vozes, pode ser utilizado para tratar *deepfakes* que violam a privacidade, ou *deepfakes* pornográficos não consensuais. O Código Penal português também pode ser aplicado neste tipo de casos, envolvendo difamação, calúnia ou injúria.

[85] INCoDe, 2019.

No que ao ordenamento jurídico espanhol diz respeito, tal como Portugal, a regulação existente é liderada pelo *EU AI Act*, com suporte de leis nacionais sobre privacidade e imagem. O Código Penal espanhol inclui disposições contra crimes que violam esses direitos, como a calúnia, a injúria e a invasão de privacidade, que podem ser usados em casos de *deepfakes*. Outras leis, como a Lei Orgânica 1/1982 sobre a Proteção Civil do Direito à Honra, Intimidade Pessoal e Familiar e à Própria Imagem, também oferecem proteção, considerando a disseminação de deepfakes sem consentimento[86].

Além das regulamentações atuais, a Espanha está a avançar com projetos legislativos específicos para abordar *deepfakes* e a Inteligência Artificial. Em outubro de 2023, foi apresentado um anteprojeto de Lei Orgânica que visa regulamentar amplamente estas matérias, propondo emendas a várias leis, incluindo a Lei 13/2022 sobre Serviços de Comunicação Audiovisual, que considera uma infração grave transmitir *deepfakes* sem consentimento, a menos que sejam claramente etiquetados, e o Código Penal, tipificando como crime o uso de *deepfakes* para prejudicar honra ou reputação. Este anteprojeto também propõe a criação de novos órgãos, como o Conselho de Participação Cidadã e o Conselho Consultivo sobre o Uso de IA, para supervisão[87]. No entanto, até março de 2025, não há indicação clara de que este projeto tenha sido aprovado.

Outro esforço significativo é o anteprojeto de Lei Orgânica para a Proteção de Menores em Ambientes Digitais, aprovado pelo Conselho de Ministros a 4 de junho de 2024. Esta iniciativa inclui medidas específicas contra *deepfakes*, como tipificar como delito a disseminação não autorizada de imagens pornográficas geradas por IA, e visa proteger menores de danos reputacionais e psicológicos causados por deepfakes[88].

No Brasil, atualmente, a regulação sobre *deepfakes* e Inteligência Artificial é limitada, o que é curioso dado o crescimento significativo e preocupante de *deepfakes* no país (830% entre 2022 e 2023)[89]. Há projetos legislativos em andamento que indicam que algumas lacunas serão preenchidas, mas, até ao momento (março de 2025), não há uma lei geral aprovada.

No entanto, é de salientar que o Tribunal Superior Eleitoral (TSE) aprovou no final de fevereiro de 2024 várias resoluções que proíbem o uso de *deepfakes* em campanhas eleitorais, em particular a Resolução TSE nº 23.732/2024[90]. Esta norma veda conteúdos sintéticos que alterem imagem ou voz de pessoas, com penalizações como cassação de mandato para candidatos que desrespeitarem as regras. Além disso, exige que qualquer uso de IA em campanhas seja claramente sinalizado, ampliando a responsabilidade de provedores redes sociais, partidos e candidatos.

Cumpre fazer ainda menção a outros projetos, como o Projeto de Lei 1272/23, que procura criminalizar *deepfakes*, e o Projeto de Lei 5.695/23, que propõe incluir *deepfakes* na Lei Maria da Penha (lei que cria mecanismos para impedir a violência doméstica e

[86] Sked, 2024.
[87] Clarke. 2023.
[88] Bustamante, 2024.
[89] Terra, 2023.
[90] Tribunal Superior Eleitoral, 2024.

familiar contra a mulher) para proteger mulheres contra a violência digital. Esses projetos ainda estão em discussão e não foram aprovados, todavia.

Também a LGPD e leis penais existentes oferecem alguma proteção, mas a rápida evolução da tecnologia exige regulamentações mais robustas.

Noutro âmbito, apesar de gerar alguma controvérsia, está em tramitação o Projeto de Lei 2630/2020[91], também conhecido como o Projeto de Lei das *Fake News,* alegado por alguns como um mecanismo de censura que se pauta pelos princípios de Liberdade, Responsabilidade e Transparência na Internet, visando regular e combater a desinformação, "desencorajando o uso de contas inautênticas". São considerados como desinformação "os conteúdos sem veracidade, manipulados, com potencial a causar danos, ficando assim vedados as contas inautênticas, disseminadores artificiais e imagens manipuladas que imitam a realidade".

Nos Estados Unidos, têm sido implementadas diversas políticas para enfrentar não só os riscos associados aos *deepfakes*, mas também a disseminação de informações falsas. Em 2019, "o Congresso aprovou o *Deepfake Report Act*, uma lei que exige que o Departamento de Segurança Interna publique relatórios anuais sobre o avanço da tecnologia *deepfake* e as suas potenciais ameaças à segurança nacional"[92]. Além disso, o Instituto Nacional de Padrões e Tecnologia lançou um programa dedicado ao desenvolvimento de padrões e diretrizes para identificar e combater notícias manipuladas.

A nível estatal também se verificam várias iniciativas para combater a tecnologia deepfake, como é o caso da Califórnia que criou uma Task Force Consultiva para o *Deepfake*, constituída por peritos jurídicos, peritos em tecnologia e representantes das comunidades académicas, sendo responsável por analisar os possíveis efeitos dos *deepfakes*, entender as consequências do seu conteúdo na sociedade e propor orientações de políticas e leis para enfrentar essa adversidade[93]. Já a lei do Estado de Virgínia contra a pornografia de vingança passou a considerar a distribuição de fotos e vídeos falsos uma infração, ampliando assim o âmbito da lei para incluir *deepfakes*.

Neste sentido, muitos autores apelam a uma regulamentação mais rigorosa dos mediam com enfase na transparência algorítmica e na definição de normas técnicas para a autenticidade dos conteúdos, propondo que agencias reguladoras trabalhem em estreita colaboração com as partes interessadas para garantir o cumprimento dessas normas. Mais uma vez se refere a necessidade de uma abordagem cooperativa e multifacetada para combater eficazmente a disseminação de *deepfakes*.

Contudo, influenciado pelo resultados das últimas eleições norte-americanas que ditaram a vitória de Donald Trump e à semelhança do que acontece com a rede social *X*, adquirida por Elon Musk, Mark Zuckerberg o proprietário da *Meta* (empresa detentora do *Facebook*, *Instagram* e *WhastApp*) anunciou no dia 7 de janeiro do presente ano o fim da

[91] Projeto de Lei 2630/2020 (3 de julho de 2020), Portal da Câmara dos Deputados. Disponível em https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735.
[92] Chawki, 2024.
[93] Ibidem em 92.

verificação de factos por especialistas, dizendo que a mesma em nada contribuía para a liberdade de expressão, sendo um mecanismo que detinha "demasiados erros e censura excessiva"[94]. Ora, em abono da verdade, tal constituiu um enorme passo atrás no combate a *fake news* e à desinformação já prevalecente.

Por fim, na ordem jurídica chinesa, país amplamente conhecido pelo seu controlo rigoroso sobre a tecnologia e internet, a regulação de *deepfakes* e Inteligência Artificial reflete uma abordagem que combina medidas específicas com leis gerais, alinhada com os objetivos de segurança nacional e estabilidade social.

A principal regulamentação específica para *deepfakes* é as "Disposições Administrativas sobre Síntese Profunda para Serviços de Informação na Internet"[95], emitida pela Administração do Ciberespaço da China (CAC) e promulgada a 25 de novembro de 2022. Estas disposições definem a síntese profunda como o uso de tecnologias com aprendizagem profunda para gerar ou modificar texto, imagens, áudios ou vídeos que possam induzir em erro. Os requisitos incluem: a "rotulagem obrigatória"**,** na qual qualquer conteúdo *deepfake* deve ser claramente rotulado para indicar que foi gerado ou modificado por tecnologia de síntese profunda; a "responsabilidades das plataformas", em que fornecedores de serviços de internet devem verificar a identidade dos seus usuários e remover conteúdos ilegais ou prejudiciais, como aqueles que violam a segurança nacional, o interesse público ou os direitos individuais; e a "avaliação de riscos**"**, sendo que as plataformas devem realizar avaliações de riscos e adotar medidas para prevenir o uso indevido da tecnologia de síntese profunda.

Para a IA em geral, a China não possui uma lei abrangente, mas há um quadro regulatório fragmentado composto por diretrizes setoriais, políticas e normas éticas - o "Plano de Desenvolvimento de Inteligência Artificial de Nova Geração"[96] de 2017, que estabelece metas para tornar a China líder global em IA até 2030, sendo, por isso, mais um plano de promoção do que de regulação propriamente dita.

Em 2022, o Ministério da Ciência e Tecnologia lançou as "Normas Éticas para a Nova Geração de Inteligência Artificial"[97], que "deve ser entendida e interpretada como parte de um sistema, e não como uma lei isolada"[98] e que fornecem diretrizes para desenvolvimento ético, mas não têm força de lei. Além disso, a Lei de Proteção de Informações Pessoais (PIPL, na sigla em inglês), promulgada em 2021, regula a recolha, o uso e o processamento de informações pessoais, aplicando-se a sistemas de IA que processam dados pessoais.

---

[94] Santos, 2025.
[95] Disposições Administrativas sobre Síntese Profunda para Serviços de Informação na Internet. Texto traduzido disponível em https://www.chinalawtranslate.com/en/deep-synthesis/. Texto original disponível em https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm.
[96] Plano de Desenvolvimento de Inteligência Artificial de Nova Geração. Disponível em https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
[97] Texto disponível em https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/.
[98] Souza, 2021.

6. Conclusão

A presente dissertação procurou analisar, de forma crítica e fundamentada, os perigos (e os desafios jurídicos) emergentes da utilização da IA Generativa no contexto do processo penal. Partindo do pressuposto de que a revolução tecnológica impôs novas formas de produção e manipulação de informação — como é o caso dos *deepfakes* —, tornou-se evidente a necessidade de reavaliar os princípios clássicos do direito processual penal à luz de um novo paradigma, profundamente marcado pela disrupção digital.

O desenvolvimento e disseminação de ferramentas baseadas em modelos de linguagem de larga escala e redes neurais profundas representam um avanço inegável em termos de eficiência, produtividade e acesso ao conhecimento. No entanto, o seu uso no domínio da justiça penal suscita legítimas preocupações quanto à integridade da prova, à proteção dos direitos fundamentais dos arguidos, à equidade processual e à preservação da autonomia decisória do juiz. Os riscos decorrentes da opacidade algorítmica, da dificuldade em verificar a origem e autenticidade de conteúdos audiovisuais, bem como do enviesamento dos dados utilizados nos sistemas preditivos, impõem uma reflexão urgente sobre os limites da intervenção tecnológica em matéria probatória e nos restantes campos da sociedade.

Verificou-se que, apesar do potencial transformador da IA, o seu uso indiscriminado pode amplificar desigualdades já existentes, colocar em causa a verdade material e comprometer o contraditório. A disparidade no acesso à perícia forense digital, a ausência de um regime uniforme sobre a autenticidade da prova gerada por IA e a inexistência de mecanismos universais de deteção de manipulações complexas reforçam a necessidade de uma abordagem multidisciplinar, regulamentada e sustentada na garantia de um julgamento justo.

Neste sentido, impõe-se a criação de instrumentos jurídicos e técnicos e adaptações à legislação em vigor que permitam compatibilizar a evolução tecnológica com os princípios fundamentais do Estado de direito democrático. Medidas como a obrigatoriedade de perícia técnica para provas audiovisuais e a instituição de modelos de perícia pública, acessíveis e imparciais, assumem-se como propostas concretas para enfrentar os desafios impostos pelos *deepfakes* e pela prova digital.

No plano normativo, iniciativas como o *EU AI Act* e a estratégia nacional AI Portugal 2030 revelam-se fundamentais, mas insuficientes enquanto não forem acompanhadas de uma verdadeira literacia digital por parte de todos: juristas e sociedade civil. A confiança no sistema de justiça depende, em última instância, da sua capacidade de se adaptar aos novos tempos sem perder de vista os valores que lhe são estruturantes: imparcialidade, verdade, proporcionalidade e igualdade.

Assim, reconhecendo-se que a IA é uma ferramenta com enorme potencial, importa garantir que o seu uso se subordine ao Direito, e não o contrário. Como bem afirmava

Martin Luther King, “a ciência pode ter encontrado a cura para quase todos os males, mas não encontrou remédio para o pior de todos: a indiferença dos homens”. A justiça penal do futuro exigirá mais do que inovação tecnológica — exigirá consciência crítica, responsabilidade institucional e compromisso ético.

7. Bibliografia

Agência Lusa (23 de dezembro de 2023). Programa com IA promete "nova era" nas decisões da gestão penitenciária. *Observador*. Disponível em https://observador.pt/2023/12/23/programa-com-ia-promete-nova-era-nas-decisoes-da-gestao-penitenciaria/.

Agência Lusa (23 de novembro de 2023). Inteligência Artificial. Juízes alertam para "riscos sérios" nos tribunais e falam em "desumanização da justiça". *Observador*. Disponível em https://observador.pt/2023/03/16/inteligencia-artificial-juizes-alertam-para-riscos-serios-nos-tribunais-e-falam-em-desumanizacao-da-justica/.

Arredondo, Pablo (19 de abril de 2023). GPT-4 Passes the Bar Exam: What That Means for Artificial Intelligence Tools in the Legal Profession. *Stanford Law School Blog*. Disponível em https://law.stanford.edu/2023/04/19/gpt-4-passes-the-bar-exam-what-that-means-for-artificial-intelligence-tools-in-the-legal-industry/.

Banh, L. & Strobel, G. (6 de dezembro de 2023). Generative artificial intelligence. *Springer Nature*. Disponível em https://link.springer.com/article/10.1007/s12525-023-00680-1.

Big Brother Watch (fev. 2020). Big Brother Watch briefing on Algorithmic Decision-Making in the Criminal Justice System. Disponível em https://bigbrotherwatch.org.uk/wp-content/uploads/2020/02/Big-Brother-Watch-Briefing-on-Algorithmic-Decision-Making-in-the-Criminal-Justice-System-February-2020.pdf.

Breen, Danielle C. (1 de janeiro de 2021). Silent no more: How deepfakes will force courts to reconsider video admission standard. *Journal of High Technology Law*. Disponível em https://law-journals-books.vlex.com/vid/silent-no-more-how-870462441.

Bustamante, M. (junho de 2024). [ES] Approval of draft Organic Law on the Protection of Minors in Digital Environments. *Iris Merlin*. Disponível em https://merlin.obs.coe.int/article/10093.

Cabral, Alvin R. (14 de maio de 2024). OpenAI's GPT-4o: What's in the new ChatGPT generative AI model and how does it work. *The National*. Disponível em https://www.thenationalnews.com/future/technology/2024/05/14/openai-chatgpt-4o/.

Carrão, M. C. (junho de 2022). Artificial Intelligence in Criminal Proceedings: The admissibility of AI-generated evidence. *Nova School of Law*, pp. 27–56. Disponível em https://run.unl.pt/handle/10362/145184.

Caetano, Edgar (18 de novembro de 2024). "Por favor, morre". Inteligência artificial da Google chama "nódoa humana" a jovem universitária. *Observador*. Disponível em https://observador.pt/2024/11/18/por-favor-morre-inteligencia-artificial-da-google-chama-nodoa-humana-a-jovem-universitaria/.

CEPEJ (2018). Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente. Disponível em https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0.

Chen, Y. et al. (13 de maio de 2023). GPT-Sentinel: Distinguishing human and ChatGPT generated content. *Cornell University*. Disponível em https://arxiv.org/abs/2305.07969.

Choi, Jonathan H. e Schwarcz, Daniel (13 de Agosto de 2023). AI Assistance in Legal Analysis: An Empirical Study. *Journal of Legal Education*. Disponível em https://ssrn.com/abstract=4539836.

Clark, O. (27 de novembro de 2023). Spain seeks to adapt its regulations to the artificial intelligence era. *Osborne Clarke*. Disponível em https://www.osborneclarke.com/insights/spain-seeks-adapt-its-regulations-artificial-intelligence-era.

Contini, Francesco, Reiling, Dory (2022). Double normalization: When procedural law is made digital. *Oñati Socio-Legal Series*, Vol. 12, nº 3, pp. 654–688. Disponível em https://opo.iisj.net/index.php/osls/article/view/1362-.

Deb, Ekata (24 de dezembro de 2023). COMPAS — an AI tool sending or keeping people in Jail. *Medium*. Disponível em https://edblogs.medium.com/compas-an-ai-tool-sending-or-keeping-people-in-jail-d9228df3a2c6.

Delfino, R. (8 de fevereiro de 2023). Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery. *Hastings Law Journal*. Disponível em https://repository.uclawsf.edu/hastings_law_journal/vol74/iss2/3/.

Delfino, R. (10 de fevereiro de 2024). Pay-to-play: Access to Justice in the Era of AI and Deepfakes. *Loyola Law School, Los Angeles Legal Studies Research Paper*. Disponível em https://ssrn.com/abstract=4722364.

D'Alessandra, F. & Sutherland, K. (7 de junho de 2021). The promise and challenges of new actors and new technologies in international justice. *Journal of International Criminal Justice*. Disponível em https://academic.oup.com/jicj/article/19/1/9/6294452.

Dietmar, J. (21 de maio de 2019). GANs and deepfakes could revolutionize the fashion industry. *Forbes*. Disponível em

https://www.forbes.com/sites/forbestechcouncil/2019/05/21/gans-and-deepfakescould-revolutionize-the-fashion-industry/#53cb7f713d17.

Diretiva (UE) 2024/1385 (14 de maio de 2024), Parlamento Europeu e do Conselho. *Jornal Oficial da União Europeia*. Disponível em https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=OJ:L_202401385.

Ettekoven, Bart Jan van & Prins, C. (28 de dezembro de 2018). Data analysis, artificial intelligence and the judiciary system. *Research Handbook in Data Science and Law*, pp. 425–447. Disponível em https://doi.org/10.4337/9781788111300.00009.

Europol (2022). Facing reality? Law enforcement and the challenge of deepfakes. *Publications Office of the European Union*. Disponível em https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf.

Fallis, D. (6 de agosto de 2020). The Epistemic Threat of Deepfakes. *Philosophy Technology*. Disponível em https://link.springer.com/article/10.1007/s13347-020-00419-2.

Finley, B. (30 de abril de 2024). Deepfake of principal's voice is the latest case of AI being used for harm. *Associated Press News*. Disponível em https://apnews.com/article/ai-maryland-principal-voice-recording-663d5bc0714a3af221392cc6f1af985e.

Floridi, L. (1 de agosto de 2018). Artificial Intelligence, Deepfakes and a Future of Ectypes. *Philos. Technol.* 31, pp. 317–321. Disponível em https://doi.org/10.1007/s13347-018-0325-3.

Freitas, Pedro Miguel (2023). Deepfakes, conteúdo gerado por inteligência artificial e verdade processual. *Em El proceso penal ante una nueva realidad tecnológica europea*, pp. 195–205. Disponível em https://dialnet.unirioja.es/servlet/articulo?codigo=8857600.

Godefroy Lêmy, Lebaron Frédéric, Levy-Vehel Jacques (julho de 2019). Comment le numérique transforme le droit et la justice vers de nouveaux usages et un bouleversement de la prise de décision. *GIP—Mission de recherche Droit et Justice*. Disponível em http://www.gip-recherche-justice.fr/publication/comment-le-numerique-transforme-le-droit-et-la-justice-par-de-nouveaux-usages-et-un-bouleversement-de-la-prise-de-decision-anticiper-les-evolutions-pour-les-accompagner-et-les-maitriser/.

Goldenstein J., Lohn, A. (23 de janeiro de 2024). Deepfakes, Elections, and Shrinking the Liar's Dividend. *Brennan Center*. Disponível em https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend.

Gong, S. E. (6 de setembro de 2024). South Korea investigates Telegram over alleged sexual deepfakes. *National Public Radio*. Disponível em https://www.npr.org/2024/09/06/nx-s1-5101891/south-korea-deepfake.

Henriques, Ana (24 de novembro de 2024). Erros crassos em acórdão levam advogados a suspeitar de inteligência artificial. *Público*. Disponível em https://www.publico.pt/2024/11/24/sociedade/noticia/erros-crassos-acordao-levam-advogados-suspeitar-inteligencia-artificial-2113160.

INCoDE (2019). Estratégia Nacional de Inteligência Artificial. Disponível em https://www.portugal.gov.pt/pt/gc21/comunicacao/documento?i=estrategia-inteligencia-artificial-2030.

John Villasenor (5 de outubro de 2024). Generative Artificial Intelligence and the Practice of Law: Impact, Opportunities, and Risks. *25 Minn. J.L. Sci. & Tech.* 25. Disponível em https://scholarship.law.umn.edu/mjlst/vol25/iss2/8.

Kietzmann, Jan, Lee, Linda, McCarthy, Ian & Kietzmann, Tim. (dezembro de 2019). *Deepfakes: Trick or treat*, Business Horizons, 63. Disponível em https://doi.org/10.1016/j.bushor.2019.11.006.

Lameira, J. S. (21-22 de outubro de 2021). Sessão de encerramento proferida pelo Vice-Presidente do CSM, Juiz Conselheiro José Sousa Lameira. *XV Encontro Anual do CSM – A (des)humanização da Justiça – Tecnologia como meio e não como fim*, Beja, Portugal.

Martins, José Eduardo. (2021). *Dilemas éticos e jurídicos do uso da inteligência artificial na prática jurídica*, RJLB, nº4, pp. 919-952. Disponível em https://www.cidp.pt/revistas/rjlb/2021/4/2021_04_0919_0952.pdf

Melville, K. (29 de agosto de 2019). *The insidious rise of deepfake porn videos — and one woman who won't be silenced*, ABC News. Disponível em https://www.abc.net.au/news/2019-08-30/deepfake-revenge-porn-noelle-martin-story-of-image-based-abuse/11437774

Mncwabe, Sammy & Schmidt, Nadine. (20 de abril de 2023). *Michael Schumacher's family planning legal action over fake AI interview*, CNN. Disponível em https://edition.cnn.com/2023/04/20/motorsport/michael-schumacher-fake-ai-interview-spt-intl/index.html

Mollick, Ethan. (14 de dezembro de 2022). *ChatGPT Is a Tipping Point for AI*, Harvard Business Review. Disponível em https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai

Murphy, H. (22 de agosto de 2024). *Transcript: The trouble with deepfakes — liar's dividend*, Financial Times. Disponível em https://www.ft.com/content/7f22ce59-1c6c-4d84-bca8-dc539992e286

Novais, Pedro & Freitas, Pedro Miguel. (30 de maio de 2018). *Inteligência Artificial e Regulação de algoritmos*, Diálogos UE-Brasil. Disponível em https://etica.uazuay.edu.ec/sites/etica.uazuay.edu.ec/files/public/49f7d3_Intelig%C3%AAncia%20Artificial%20e%20Regula%C3%A7%C3%A3o%20de%20Algoritmos.pdf

Parreira, Rui. (15 de maio de 2025). *GPT-4.1 já está disponível para utilizadores do ChatGPT Plus, Pro e Team*, SapoTek. Disponível em https://tek.sapo.pt/noticias/computadores/artigos/gpt-4-1-ja-esta-disponivel-para-utilizadores-do-chatgpt-plus-pro-e-team

Pedroso, João & Santos, Andreia. (agosto de 2024). *Inteligência artificial e justiça criminal: Riscos e desafios*, Revista SOCIOLOGIA ON LINE, nº 35, pp. 134-155. Disponível em https://revista.aps.pt/pt/inteligencia-artificial-e-justica-criminal-riscos-e-desafios/

Pfefferkorn, R. (1 de outubro de 2020). *Deepfakes in the Courtroom*, Boston University Public Interest Law Journal, Vol. 29, No. 2. Disponível em https://ssrn.com/abstract=4321140

Reiling, A. D. (Dory). (24 de novembro de 2020). *Courts and Artificial Intelligence*, International Journal for Court Administration. Disponível em https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3736411

Rodrigues, A. M. (2023). *Justiça penal e inteligência artificial – uma justiça fitness?*, em C. Aranguena Fanego et al. (Dirs.), *El proceso penal ante una nueva realidad tecnológica europea*, pp. 207–230. Disponível em https://dialnet.unirioja.es/servlet/articulo?codigo=8857599

Sako, Mari. (6 de março de 2024). *How Generative AI Fits into Knowledge Work*, Communications of the ACM. Disponível em https://cacm.acm.org/opinion/how-generative-ai-fits-into-knowledge-work/

Santos, C. (8 de janeiro de 2025). *"Check". Meta dá passo atrás no combate a "fake news" e desinformação*, RTP Notícias. Disponível em https://www.rtp.pt/noticias/mundo/check-meta-da-passo-atras-no-combate-a-fake-news-e-desinformacao_n1625836

Santos, Gerardo. (12 de novembro de 2024). *Primeiro-ministro anuncia lançamento de modelo de linguagem de IA em português*, Diário de Notícias. Disponível em https://www.dn.pt/2595648740/primeiro-ministro-anuncia-lancamento-de-modelo-de-linguagem-de-ia-em-portugues

Sandoval, MP., de Almeida Vau, M., Solaas, J. & Rodrigues, L. (2024). *Threat of deepfakes to the criminal justice system: a systematic review*, Crime Sci 13. Disponível em https://doi.org/10.1186/s40163-024-00239-1

Sepúlveda, Darío Parra & Machuca, Ricardo Concha. (29 de outubro de 2021). *Inteligencia artificial y derecho: Problemas, desafíos y oportunidades*, 70 Vniversitas. Disponível em https://doi.org/10.11144/Javeriana.vj70.iadp

Silva, Germano Marques da. (2008). *Curso de Processo Penal*, Vol. II, 4ª ed., Editorial Verbo.

Sked, A. (7 de novembro de 2024). *Crimes against honour, privacy and one's own image in Spain*, Conesa Legal. Disponível em https://www.conesalegal.com/en/info/crimes-against-honour-privacy-and-ones-own-image-in-spain

Souza, C. (22 de novembro de 2021). *A Lei de Proteção de Informações Pessoais (PIPL) e o Papel do Direito numa China Hiperconectada*, Observa China. Disponível em https://www.observachina.org/pt/articles/a-lei-de-protecao-de-informacoes-pessoais-pipl-e-o-papel-do-direito-numa-china-hiperconectada

Terra. (28 de novembro de 2023). *Deepfakes crescem 830% no Brasil em um ano, aponta estudo*. Disponível em https://www.terra.com.br/noticias/deepfakes-crescem-830-no-brasil-em-um-ano-aponta-estudo%2C601f3d28caa943b3728390a82f13cc2b0x3pdhbk.html

Tribunal Superior Eleitoral. (28 de fevereiro de 2024). *TSE proíbe uso de inteligência artificial para criar e propagar conteúdos falsos nas eleições*. Disponível em https://www.tse.jus.br/comunicacao/noticias/2024/Fevereiro/tse-proibe-uso-de-inteligencia-artificial-para-criar-e-propagar-conteudos-falsos-nas-eleicoes

Verdovila, L. (18 de janeiro de 2020). *Media Forensics and DeepFakes: an overview*, Cornell University. Disponível em https://arxiv.org/abs/2001.06564

Wall, Matthew (5 de julho de 2019). *Inteligência artificial: Por que as tecnologias de reconhecimento facial são tão contestadas*, em BBC News Brasil. Disponível em https://www.bbc.com/portuguese/geral-48889883. Consultado a 1 de junho de 2025.

Westerlund, M. (novembro de 2019). *The emergence of Deepfake technology: A review*, Technology Innovation Management Review, pp. 39–52. Disponível em https://timreview.ca/article/1282

Xu, Ziwei, Jain, Sanjay & Kankanhalli, Mohan. (22 de janeiro de 2024). *Hallucination is Inevitable: An Innate Limitation of Large Language Models*, Cornell University. Disponível em https://arxiv.org/pdf/2401.11817

Zahn, Max. (9 de dezembro de 2022). *What is ChatGPT, the artificial intelligence text bot that went viral?* ABC News. Disponível em https://abcnews.go.com/Technology/chatgpt-artificial-intelligence-text-bot-viral/story?id=94857599

Zambon Perícia. (20 de julho de 2021). *Entenda as diferenças da atuação da perícia judicial no Brasil e nos EUA*. Disponível em https://zambonpericia.com.br/entenda-as-diferencas-da-atuacao-da-pericia-judicial-no-brasil-e-nos-eua/