

Capítulo 6

Deepfakes, conteúdo gerado por inteligência artificial e verdade processual

PEDRO MIGUEL FREITAS¹

Doutor em Direito. Professor da Escola do Porto da Faculdade de Direito Universidade Católica Portuguesa

I. INTRODUÇÃO

A procura da verdade histórica no decurso de um processo judicial, em particular o penal, é um caminho especialmente tortuoso e repleto de dificuldades. Desde logo, as colocadas pelas limitações intrínsecas à cognição humana dos factos que constituem o objeto do processo, que se somam às que resultam da circunstância de o decisor judicial não ter, em regra, contactado diretamente com a situação que deve julgar. Porque o decisor judicial não presenciou a comissão ou omissão da conduta penalmente relevante, o contacto com os factos ocorre de forma mediata, por intermédio de documentos ou testemunhos, por exemplo. Contacto esse que, mesmo que fosse direto, não é impermeável, antes pelo contrário, à subjetividade da percepção humana e ao circunstancialismo objetivo do decurso da situação apreciada.

Naturalmente que o princípio da imediação não resolve este problema. Tem o seu âmbito de aplicação confinado ao processo penal em si, não retrocedendo, por razões claras para todos, até ao momento da consumação típica do crime. A sua influência faz-se sentir, isso sim, na fase de julgamento. O motivo?

1. O presente texto serviu de suporte a uma comunicação oral apresentada na Conferência Ibero-Atlântica sobre Justiça Penal e Novas Tecnologias, que teve lugar nos Açores, em 13 de junho de 2022. Optou-se por manter o registo simples e tendencialmente coloquial do texto de apresentação.

Antes de mais, é preciso atender à natureza deste princípio. Alcandorando-se em princípio fundamental do processo penal, o seu conteúdo nutre-se das linhas caracterizadoras do direito processual penal positivado² e ao mesmo tempo serve de critério para integração de lacunas desse mesmo direito positivo. Assim, o princípio da imediação cinge-se ao tempo do processo penal e, dentro deste, essencialmente à audiência de julgamento.

Quanto ao seu conteúdo, convém relembrar que, tal como os princípios da publicidade e o da oralidade, é um princípio que se relaciona com a forma dos atos processuais e implica que “a decisão jurisdicional só [possa] (...) ser proferida por quem tenha assistido à produção das provas e à discussão da causa pela acusação e pela defesa” e que se deva “dar preferência aos meios de prova que se encontrem em relação mais directa com os factos probandos (v.g. preferência das testemunhas presenciais às de «ouvir dizer», dos documentos originais às suas cópias, etc.)”³.

A imediação “torna possível, na apreciação das provas, a formação de um juízo insubstituível sobre a credibilidade da prova; das razões que se podem observar, no exame directo da prova, para acreditar, ou não acreditar, na mesma” (Ac. STJ de 06/02/2008, processo n.º 07P4374).

Mesmo assim configurado, o princípio da imediação não resolve definitivamente o problema da aproximação –pois não é mais do que isso– à verdade histórica ou ontológica. A imediação processual mais não é que um contato direto e pessoal entre o juiz e os meios de prova⁴, uma espécie de aproximação mediada ao facto ilícito típico, que não se confunde com a imediação pré-processual ou extraprocessual associada a uma exposição presencial ao mesmo.

II. INTELIGÊNCIA ARTIFICIAL E DIREITO PENAL

Esta aproximação enfrenta agora um outro desafio colocado pelas novas tecnologias. Os avanços sentidos no campo da inteligência artificial, nomeadamente em *machine learning*, têm sido revolucionários, ao mesmo tempo que demonstrativos de uma potencialidade que promete transformar profundamente o rumo da civilização humana. Enquanto isso, ao Direito, enquanto instrumento regulatório, é-lhe pedido que normativize esta realidade, marcando a fronteira do lícito e do ilícito, do desejável e do indesejável, do risco permitido e do proibido.

2. SILVA, Germano Marques, *Curso de Processo Penal*, vol. 1, 6.ª ed., Verbo, 2010, p. 64.

3. SILVA, Germano Marques, *Curso de Processo Penal*, vol. 1, 6.ª ed., Verbo, 2010, p. 105.

4. SILVA, Germano Marques, *Direito Processual Penal Português, Do Procedimento (Marcha do Processo)*, vol. 3, Universidade Católica Editora, 2015, p. 212.

Mas desta vez não se trata somente de incorporar no *corpus iuris* algo que lhe é estranho e externo.

Tomemos como exemplo o direito penal em sentido amplo. A emergência e desenvolvimento de técnicas de inteligência artificial tem contribuído para a criação de novos modos de preenchimento típico de comportamentos que convençamos dizer que são pertencentes à criminalidade tradicional (p. ex. difamação) e fazem-nos refletir sobre a eventual necessidade de identificação de bens jurídico-penais até aqui inexistentes na tessitura jurídico-criminal como resposta aos mais recentes ensejos comunitários no sentido da proteção de estados, objetos ou bens de relevância individual ou coletiva. Dito de outro modo, o condão transformativo da inteligência artificial sente-se e sentir-se-á cada vez mais intensamente nos mais ínfimos aspetos da nossa vida, incluindo, por maioria de razão, aqueles com que o direito penal substantivo terá de lidar, ou por via da neocriminalização ou pela acomodação de modos de execução inauditos, se não mesmo pelo reequacionamento das suas categorias dogmáticas e princípios norteadores.

Atenta a sua simbiótica relação com o direito penal substantivo, ao processo penal é exigido idêntico esforço de reflexão e eventual adaptação a uma nova realidade. Neste domínio, que se cruza com o anterior, não são incomuns os estudos acerca do impacto jusfundamental de ferramentas algorítmicas de suporte à decisão judicial. Em particular, como o seu funcionamento, tendencialmente opaco e permeável a vieses sociais, mina o direito a um julgamento justo e impossibilita o exercício efetivo de garantias de defesa como o direito ao recurso. Outro dos direitos fundamentais potencialmente afetado pelo uso de IA é o direito à privacidade. A crescente presença de equipamentos inteligentes (*smart devices*), p. ex. colunas inteligentes, que vão registando e conservando o que é dito ao seu redor origina situações em que a “AI in some cases can actually assume the role of a witness, providing crucial evidence in an investigation or lawsuit”⁵, no sentido da condenação ou da absolvição. Tudo isto com consequências a nível probatório. Uma outra questão, sobre a qual nos dedicaremos um pouco mais, é a da criação ou manipulação de fotos, vídeos e áudio com recurso a inteligência artificial e seu impacto no processo penal.

III. DEEPFAKES E O DIREITO PROBATÓRIO

A manipulação de imagens (estáticas ou em movimento) e áudio não é um fenómeno exclusivo do século XXI. O seu aparecimento

5. BARFIELD, Woodrow e Ugo Pagallo, *Advanced Introduction to Law and Artificial Intelligence*, Elgar, 2020, p. 51.

entrelaça-se com a história da gravação do áudio, fotografia ou vídeo⁶, inclusivamente quando o formato digital ainda não existia. Aliás, muito antes sequer do advento do Photoshop e pouco tempo após a invenção da fotografia, Hippolyte Bayard criou aquela que é considerada a primeira fotografia falsa, intitulada “Noyé” (1840), um autorretrato de Boyard supostamente afogado após suicídio, como forma de protesto contra a falta de reconhecimento público do seu trabalho na invenção da fotografia⁷. Nesse mesmo século XIX surgiram as fotografias espíritas onde se empregava uma técnica de dupla exposição com a qual se sobrepunham várias imagens numa só fotografia e que supostamente revelariam fantasmas.

Desde então, a manipulação de fotografias foi-se tornando trivial. Apesar da sua falibilidade, a fotografia foi sendo admitida enquanto meio de prova sujeito à livre apreciação do julgador. No Reino Unido, por exemplo, material fotográfico como prova incriminatória do arguido é admissível em quatro hipóteses ([2003] 1 Cr APP Rep 21):

“(i) quando a imagem fotográfica for suficientemente nítida, o júri pode compará-la com o arguido sentado no banco dos réus;

(ii) quando uma testemunha conhece o arguido suficientemente bem para reconhecê-lo como o infrator retratado na imagem fotográfica, ele pode depor sobre isso; e isso pode acontecer mesmo que a imagem fotográfica não esteja mais disponível para o júri;

(iii) quando uma testemunha que não conhece o arguido passa muito tempo a ver e a analisar imagens fotográficas do local, adquirindo assim conhecimentos especiais que o júri não possui, pode fornecer prova de identificação com base na comparação entre essas imagens e uma fotografia razoavelmente contemporânea do arguido, desde que as imagens e a fotografia estejam à disposição do júri;

(iv) um especialista devidamente qualificado com habilidades de mapeamento facial pode fornecer provas de identificação com base numa comparação entre imagens da cena (aperfeiçoadas ou não e uma fotografia razoavelmente contemporânea do arguido, desde que as imagens e a fotografia estejam disponíveis para o júri”.

Poder-se-á perguntar o que traz de novo a inteligência artificial. O que há de diferente na criação e manipulação de imagens e áudio por

6. KIETZMANN, Jan, Linda W. Lee, Ian P. McCarthy, Tim C. Kietzmann, “Deepfakes: Trick or treat?”, in *Business Horizons*, volume 63, número 2, 2020, p. 145.

7. WHEELER, Thomas H., *Phototruth or Photofiction?: Ethics and Media Imagery in the Digital Age*, Routledge, 2002, p. 15.

intermédio da inteligência artificial? Sumariamente: a credibilidade e a acessibilidade⁸.

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. Algo similar existia já na área do cinema –efeitos especiais e visuais–, mas requerem mão-de-obra especializada e intensiva, bem como hardware e software adequados.

A credibilidade inerentemente atribuída a algo que vemos ou ouvimos é precisamente o vetor explorado por quem se serve destas novas técnicas. Pense-se, por exemplo, no fabrico de um vídeo em que certo indivíduo aparece a pontapear outrem, tendo isto servido para corroborar a acusação.

Este novo fenómeno é denominado *deepfake*. Nas palavras de der Sloot e Wagenveld, um *deepfake* é “conteúdo (vídeo, áudio ou outro) que é total ou parcialmente fabricado ou conteúdo existente (vídeo, áudio ou outro) que tenha sido manipulado”⁹.

De um ponto de vista tecnológico, a criação de *deepfakes* faz-se, em regra, com Redes Adversárias Generativas (*Generative Adversarial Networks* –GAN). Estas redes são uma espécie de algoritmo generativo que consiste na implementação de dois modelos computacionais que competem entre si: um gerador e um discriminador. A função do primeiro modelo é a de “criar exemplos que são o mais realistas possíveis para enganar o discriminador, enquanto o discriminador tenta distinguir os exemplos falsos gerados dos exemplos verdadeiros”¹⁰. Desta competição entre os dois modelos – tipicamente redes neuronais artificiais¹¹ – resulta uma aprendizagem automatizada útil para aplicações tão diversas quanto a manipulação e síntese de imagens, incluindo a criação de imagens de rostos ou objetos, a síntese de texturas, a deteção de objetos, geração de música e

8. KIETZMANN, Jan, Linda W. Lee, Ian P. McCarthy, Tim C. Kietzmann, “Deepfakes: Trick or treat?”, in *Business Horizons*, volume 63, número 2, 2020, pp. 136-137.

9. SLOOT, Bart van der, Yvette Wagenveld, “Deepfakes: regulatory challenges for the synthetic society”, in *Computer Law & Security Review*, volume 46, setembro 2022, p. 1.

10. GUI, Jie, Zhenan Sun, Yonggang Wen, Dacheng Tao, Jieping Ye, “A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications”, in *arXiv*, 2020, p. 2.

11. GOODFELLOW, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, in *arXiv*, 2014, p. 140.

vídeo, síntese de fala, processamento de linguagem natural e a descoberta de medicamentos¹², para citar apenas algumas das áreas que beneficiaram desta técnica, que é estudada deste 2014¹³.

Ademais, é uma técnica que dispensa supervisão humana para atingir os objetivos propostos. Daí que se categorize como uma modalidade de “aprendizagem não supervisionada”. Se se tratasse de uma “aprendizagem supervisionada”, haveria um programador/utilizador humano que “ensinaria” o algoritmo a partir de um conjunto de dados. Ao algoritmo é “fornecido um conjunto de dados de pares de exemplos de entrada e de saída” e “aprendem a associar cada entrada a uma saída”¹⁴, p. ex., associar uma foto de um cão a este tipo de animal e não a um gato.

Uma vez que a aprendizagem supervisionada é intrinsecamente dependente da intervenção humana e suas capacidades e limitações, assim como de um número considerável de dados, não raras vezes na ordem dos milhões, com a aprendizagem não supervisionada ultrapassam-se as referidas desvantagens¹⁵.

Os *deepfakes* podem ser agrupados em cinco tipos¹⁶: substituição facial, reencenação facial, geração de rostos, síntese de fala e falsificações superficiais (*shallowfakes*). Com a substituição facial transfere-se a imagem da face de uma pessoa para a de outra pessoa. A reencenação facial consiste em usar a imagem de alguém e recriar movimentos faciais, incluindo os lábios. Uma possível utilidade deste tipo de *deepfakes* é a de simular que alguém está a concordar ou discordar de algo ou ainda que está a dizer algo que na realidade não está. Os *deepfakes* podem também corresponder a rostos de pessoas inexistentes. Nesta modalidade de *deepfakes*, são empregadas técnicas de inteligência artificial para gerar “fotografias” de rostos (p. ex. <https://generated.photos/faces>). Com a síntese de fala é

12. HONG, Yongjun, Uiwon Hwang, Jaeyoon Yoo, Sungroh Yoon, “How Generative Adversarial Networks and Their Variants Work: An Overview”, in *arXiv*, 2019, pp. 1-2.

13. GUI, Jie, Zhenan Sun, Yonggang Wen, Dacheng Tao, Jieping Ye, “A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications”, in *arXiv*, 2020, p. 1.

14. GOODFELLOW, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, in *arXiv*, 2014, p. 139.

15. GOODFELLOW, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, in *arXiv*, 2014, p. 139.

16. TREND MICRO RESEARCH, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), *Malicious Uses and Abuses of Artificial Intelligence*, Trend Micro Research, 2020, pp. 53-54.

possível treinar algoritmos com a voz de alguém para, posteriormente, ser criada ou modificada uma comunicação oral. Finalmente, por falsificações superficiais entendem-se as falsificações de menor complexidade, em que são usadas “técnicas de edição rudimentar”¹⁷.

Neste ano de 2022, foram disponibilizadas ao público em geral ferramentas de text-to-image (texto para imagem) que, beneficiando dos avanços do *deep learning*, permitem aos seus utilizadores fornecer uma descrição textual do resultado que procuram obter e o sistema de IA cria a imagem correspondente. Ou seja, ferramentas como o DALL-E 2 (<https://openai.com/dall-e-2/>), Stable Diffusion (<https://stability.ai/>) ou o Midjourney (<https://www.midjourney.com/>) são “sistema[s] de IA que conseguem criar imagens realistas e arte a partir de uma descrição em linguagem natural” (<https://openai.com/dall-e-2/>). Em termos sintéticos, trata-se de síntese text-to-image (texto para imagem). Estas ferramentas integram também funcionalidades como o *inpainting* (criação ou alteração de partes de uma imagem sintética ou real), *outpainting* (extensão de uma imagem previamente existente) e *image-to-image* (criação de uma imagem a partir de uma imagem e uma descrição textual previamente fornecidos pelo utilizador).

Diante deste estado de coisas e, sobretudo, do que está para vir, não será exagero afirmar-se que as “tecnologias digitais parecem minar a nossa confiança acerca da natureza autêntica, genuína e original do que vemos e ouvimos”¹⁸. Inclusive quando um julgador tem a tarefa de descortinar a verdade histórica a partir dos contributos probatórios que encontra ou que traz para os autos de um processo. Como alertam 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. Isso criaria novos obstáculos legais para investigadores e advogados, além de prejudicar a credibilidade dos processos, incluindo as instituições e indivíduos que participam do mesmo processo. Até mesmo a administração ou o sistema de justiça podem ser questionados”¹⁹.

-
17. TREND MICRO RESEARCH, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), *Malicious Uses and Abuses of Artificial Intelligence*, Trend Micro Research, 2020, p. 54.
 18. FLORIDI, Luciano, “Artificial Intelligence, Deepfakes and a Future of Ectypes”, in *Philosophy & Technology*, volume 31, 2018, p. 320.
 19. TREND MICRO RESEARCH, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), *Malicious Uses and Abuses of Artificial Intelligence*, Trend Micro Research, 2020, p. 61.

Uma resposta adequada aos riscos criados pelos *deepfakes* é necessariamente multidisciplinar: “(...) [o] que o digital quebra ele também pode reparar, não muito diferente da luta interminável entre vírus e anti-vírus. No nosso caso, para além de educarmos as pessoas, adquirir novas sensibilidades e ter o enquadramento legal adequado”²⁰. São essenciais a educação e a literacia digital das pessoas, de modo a que a acessibilidade às novas tecnologias seja acompanhada de uma atitude mais crítica e reflexiva acerca da informação que é consumida por seu intermédio. A preparação da comunidade em geral para a realidade da má utilização de *deepfakes* é importante, mas a tecnologia em si também pode ser útil para minorar este problema. Por exemplo, ferramentas automatizadas de deteção de *deepfakes* ou a verificação de origem de um conteúdo audiovisual com recurso a *blockchains*²¹.

No que ao direito processual penal se refere, não restam dúvidas de que os *deepfakes* podem contaminar o acervo probatório e a decisão final do julgador. Relembre-se a hipótese acima referida de o assistente fabricar um vídeo em que é alegadamente agredido pelo arguido, ou imagine-se que são submetidas como provas vídeos ou fotos adulterados ou artificialmente criados. Atendendo ao quadro normativo português atual, sobressai desde logo o princípio da livre apreciação da prova. Previsto no artigo 127.º do Código de Processo Penal (CPP), estabelece que “a prova é apreciada segundo as regras da experiência e a livre convicção da entidade competente”. Princípio esse que comporta uma dimensão positiva e uma dimensão negativa: “significa, negativamente, a ausência de critérios legais que predeterminem o valor da prova e, positivamente, que as entidades a quem caiba valorar a prova o façam de acordo com o dever de perseguir a realização da justiça e a descoberta da verdade material, numa apreciação que terá de ser sempre objetivável, motivável e, por conseguinte, suscetível de controlo”²². O juiz de julgamento, o juiz de instrução e o ministério público devem então valorar a prova não de forma predeterminada pela lei – a isto corresponderia o princípio da prova legal – mas de acordo com as regras de experiência e a sua livre convicção. Não é, todavia, um princípio sem exceções, como evidenciam p. ex. as regras em matéria de depoimento indireto (artigo 129.º do CPP) ou de vozes públicas e convicções pessoais (artigo 130.º do CPP). Vale, porém, sem especiais limitações para a apreciação de reproduções fotográficas, cinematográficas ou fonográficas.

20. FLORIDI, Luciano, “Artificial Intelligence, Deepfakes and a Future of Ectypes”, in *Philosophy & Technology*, volume 31, 2018, p. 320.

21. TREND MICRO RESEARCH, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), *Malicious Uses and Abuses of Artificial Intelligence*, Trend Micro Research, 2020, p. 62.

22. ANTUNES, Maria João, *Direito Processual Penal*, 2.ª edição, Almedina, 2018, p. 177.

Tendo como pano de fundo o princípio da livre apreciação da prova, impõe-se perguntar se o surgimento dos *deepfakes* demandam uma reavaliação dos atuais princípios norteadores do direito processual penal português e, concomitantemente, do direito positivo.

A manutenção do *status quo* normativo –de sujeição à livre apreciação da prova– encontra um inconveniente que é o do relativo, quando não mesmo generalizado, desconhecimento desta nova realidade. Como pode ser materializado jurídico-empiricamente o princípio da livre apreciação da prova nos dias atuais se a mundividência de quem procede à valoração for sustentada em pressupostos datados, próprios de quem equivoca passado, presente e futuro, e desconhece a atualidade dos perigos para a comunidade e para a administração da justiça em particular? 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.

Uma via de encerramento das portas do sistema de justiça penal à sua instrumentalização pelos *deepfakes* passaria pela introdução de regras de admissibilidade de provas fotográficas, cinematográficas ou fonográficas com as quais se atribuiria um ónus de comprovação de autenticidade ao sujeito processual que as pretende juntar ao processo. Particularmente se se tratar do assistente ou do arguido. Desse modo, incumbiria ao sujeito processual juntar um relatório pericial atestando que, à luz dos conhecimentos tecnológicos atuais, não teria sido possível detetar indícios de que esse conteúdo fosse parcial ou totalmente artificial. A eventual crítica de parcialidade na redação do referido relatório técnico poderia ser ultrapassada com a atribuição deste encargo a um estabelecimento, laboratório ou serviço oficial, como acontece de resto com qualquer prova pericial nos termos do CPP atualmente em vigor (artigo 152.º).

A vantagem de uma solução assim configurada seria a de conferir maior confiança e tranquilidade à entidade que devesse valorar uma prova desta natureza. Embora pudesse constituir uma desvantagem naquelas situações em que a confiança e tranquilidade dessa entidade assentassem num falso negativo quanto à identificação do *deepfake*. Sairia exacerbado o potencial lesivo do *deepfake* agora que ultrapassado e ludibriado o crivo da prova pericial. Além de que se colocaria a questão de saber quem deveria arcar com as despesas inerentes à contratação de um perito. Recaindo, como seria expectável, sobre quem quisesse juntar ao processo a prova, não seria um desincentivo à sua utilização? Não estaríamos assim a caminhar para um processo desprovido de fotografias, vídeos e áudios enquanto material probatório? Ou mesmo a cavar um

fosso probatório entre sujeitos processuais consoante a sua capacidade financeira?

Provavelmente a resposta mais direta e imediata aos *deepfakes* passaria pela produção de prova pericial nos termos do artigo 151.º e ss. do CPP. Dada a complexidade técnica inerente ao juízo de atribuição de uma origem sintética a um conteúdo visual ou sonoro, o recurso à prova pericial é facilmente percebido como uma solução evidente para estes casos. A intervenção de um perito justifica-se pela circunstância de serem precisos especiais conhecimentos técnicos que o jurista não dispõe. Por essa razão, temos a presunção de que o juízo técnico esteja subtraído à livre apreciação do julgador (artigo 163.º, n.º 1 do CPP). Todavia, a questão que se deve colocar é a seguinte: se a constatação da necessidade da perícia é um elemento constitutivo do despacho que a ordena, é expectável que o julgador admita a possibilidade de que o que vê com os seus olhos e ouve com os seus ouvidos não é real?

O juízo de ponderação acerca da necessidade da prova pericial não é verdadeiramente livre e esclarecido se desconhece os pressupostos (tecnológicos) que servem de fundamento ao requerimento. Atribuir a solução de um problema complexo a um especialista, pressupõe, em primeiro lugar e antes de mais, a identificação da complexidade do problema.

E este problema poderá assumir proporções especialmente graves quando o alvo destes conteúdos sintéticos ou artificiais seja uma pessoa particularmente indefesa, alguém incapaz de, por si mesmo, clamar por justiça perante uma prova que não corresponde à realidade.

IV. CONCLUSÃO

A inteligência artificial tem fomentado mudanças galopantes na nossa sociedade, relativamente às quais o direito não pode ficar indiferente. Ao direito é demandada uma resposta normativa, enquanto é, concomitantemente, cada vez mais afetado e transformado, no desenvolvimento das suas tarefas, pelos contributos produzidos no domínio da inteligência artificial.

O surgimento dos *deepfakes*, intrinsecamente ligado a sistemas de inteligência artificial, é mais um (novo) desafio colocado ao direito penal, nomeadamente adjetivo, podendo inclusive originar uma reavaliação dos seus atuais princípios norteadores. De entre eles, merece particular atenção o da livre apreciação da prova. A questão é sobretudo a de saber se a mera existência de *deepfakes*, e risco associado de perda de confiança

na veracidade e autenticidade de factos penalmente relevantes exibidos em conteúdos audiovisuais, implicaria, *ab initio*, a subtração deste tipo de material probatório à livre apreciação do julgador.

Tendencialmente poder-se-á afirmar que a resposta deverá implicar uma maior utilização da prova pericial, desde que capaz de aferir, para além de qualquer dúvida significativa, a origem sintética ou não de um conteúdo audiovisual. No entanto, uma resposta mais holística não deve ignorar a importância da literacia digital, em particular daqueles que integram o sistema de justiça penal.