

INTELIGÊNCIA ARTIFICIAL E JUSTIÇA CRIMINAL

Riscos e desafios

ARTIFICIAL INTELLIGENCE AND CRIMINAL JUSTICE

Risks and challenges

João Pedroso

FUNÇÕES: Concetualização, Investigação, Redação do rascunho original, Redação – revisão e edição
AFILIAÇÃO: Faculdade de Economia e Centro de Estudos Sociais da Universidade de Coimbra. Av. Dr. Dias da Silva 165, 3004-512 Coimbra, Portugal
E-mail: jpedroso@fe.uc.pt | ORCID: <https://orcid.org/0000-0001-8956-2250>

Andreia Santos

FUNÇÕES: Concetualização, Investigação, Redação do rascunho original, Redação – revisão e edição
AFILIAÇÃO: Investigadora independente. 3030-250 Coimbra, Portugal
E-mail: andreiasant1@hotmail.com | ORCID: <https://orcid.org/0000-0003-2569-6991>

Resumo: A inteligência artificial (IA) veio alavancar a digitalização na área da justiça. A inovação promovida pela IA trouxe novas vantagens, sobretudo, na análise e tratamento dos dados, porém, trouxe igualmente novos desafios ao direito e sua aplicação. Entre eles, destaca-se o uso da IA na área criminal, sob a forma de uma “justiça preditiva”. Partindo de experiências no âmbito europeu e internacional com aplicação de programas de justiça preditiva com recurso à IA, debate-se os riscos e desafios face ao respeito pelos direitos fundamentais na aplicação desta nova forma de justiça. Esta análise deixa claro que a utilização de algoritmos nas decisões judiciais criminais através da justiça preditiva pode alterar o carácter dos dados e resultar em enviesamentos discriminatórios com impacto nos resultados. Além disso, ao serem vistos como objetivamente imparciais, esses algoritmos podem influenciar a independência dos juízes nas suas decisões, comprometendo assim o direito a um julgamento justo. A concluir, sugerem-se alguns elementos que podem contribuir para uma agenda futura de reflexão e de análise sobre a utilização da IA na justiça criminal e a sua conformidade (ou não) com os direitos fundamentais.

Palavras-chave: inteligência artificial, direito, área criminal, justiça preditiva.

Abstract: Artificial intelligence (AI) has boosted digitalization in the judiciary. The innovation promoted by AI has brought new advantages, especially in the analysis and processing of data, but it has also brought new challenges to the law and its application. Among these, the use of AI in the criminal area stands out, in the form of “predictive justice”. Based on experiences in the European and international spheres with the application of predictive justice programs using AI, we discuss the risks and challenges regarding respect for fundamental rights in the application of this new form of justice. This study makes it clear that the use of algorithms in criminal judicial decisions through predictive justice can alter the nature of the data and result in discriminatory biases that impact the outcomes. Moreover, by being perceived as objectively impartial, these algorithms can influence judges’ independence in their decisions, thereby compromising the right to a fair trial. To conclude, we suggest some elements that can contribute to a future agenda of reflection and analysis on the use of AI in criminal justice and its compliance (or not) with fundamental rights.

Keywords: artificial intelligence, law; criminal field, predictive justice.

Introdução

O direito e a administração da justiça têm estado sujeitos a uma transformação acelerada, pelo que os estudos sobre a política pública e as reformas da administração da justiça, nas últimas décadas, têm focado inevitavelmente a questão da inovação tecnológica (Pedroso, 2002, 2006; Pedroso & Dias, 2004). Desde o final do século XX, as novas tecnologias de informação e comunicação apresentam um enorme potencial de transformação do sistema judicial, tanto na administração e gestão da justiça, na transformação do exercício das profissões jurídicas, como na democratização do acesso ao direito e à justiça (Santos, 2005). A linha evolutiva padrão na maioria dos países europeus, remonta aos anos 1980, com a estruturação de bases de dados, e já na década de 1990, grande parte desses países desenvolveu recursos de tecnologias de informação e comunicação (Fabri & Contini, 2001). O uso de tecnologias de informação nos tribunais advém de um processo contínuo (Reiling, 2009), cuja principal mudança nos dias de hoje assenta sobre o potencial da digitalização, a qual está a ser consideravelmente ampliada pelos avanços da inteligência artificial (IA)¹. Num primeiro momento, o debate centrou-se no uso de novas tecnologias na gestão judicial, e atualmente assenta no modo como a digitalização com recurso à IA tem impacto sobre a eficiência e qualidade da justiça, bem como a sua influência na aplicação do direito (Biard et al., 2021; Donoghue, 2017; Wallace, 2017).

A adoção de novas tecnologias de informação e comunicação com recurso à IA tem sido uma das principais estratégias com o objetivo de melhorar a administração judicial de vários países, realçando a sua natureza instrumental (Contini, 2020). A IA tem sido utilizada enquanto ferramenta na gestão judicial; no auxílio à decisão judicial; no acesso ao direito e à justiça através da disponibilização de informação e da resolução de litígios *online*; e na segurança pública e investigação criminal. O seu impacto na gestão judicial e ampliação do acesso à justiça, nomeadamente, nos tribunais, tem operado uma transformação da sua organização e gestão através da reengenharia de procedimentos e processos de trabalho, modificando a natureza do trabalho, reconfigurando as profissões jurídicas e a prestação do serviço público de Justiça (Pedroso, 2021; Sourdin, 2018). Alguns autores, com

1 Segundo a definição consensualizada no Artificial Intelligence Act (AI ACT, 2024, p. 166), um sistema de IA é:

- um sistema baseado em máquinas concebido para funcionar com níveis de autonomia variáveis, que pode apresentar capacidade de adaptação após a implantação e que, para objetivos explícitos ou implícitos, e com base nos dados de entrada que recebe, infere a forma de gerar resultados, tais como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais.

base na tecnologia que está a ser utilizada por tribunais e sistemas judiciais para fornecer suporte, admissão e processos consultivos que visam ajudar os litigantes a negociar de forma mais eficaz, sem envolver funcionários do tribunal ou outros profissionais, preveem que muitos tribunais continuarão a construir e ampliar plataformas e sistemas *online* que suportam arquivamento, encaminhamento e outras atividades (Sourdin, 2018). A título ilustrativo, a recente crise sanitária despontada pela pandemia da COVID-19, em 2020, serviu como catalisador do uso da IA, com os tribunais a verem-se forçados a prestar os seus serviços quase exclusivamente *online*, de que é exemplo a proliferação da “Online Dispute Resolution” (ODR) – resolução alternativa de litígios *online* (Brito & Fernandes, 2020; Thompson, 2015).

A inovação promovida pela IA trouxe novas vantagens, sobretudo, na análise e tratamento dos dados, porém, trouxe igualmente novos desafios ao direito e sua aplicação. (Campbell, 2020; Contini, 2020; Reiling, 2020). No âmbito europeu, e partindo de experiências quanto ao uso de programas com recurso à IA na justiça, o foco centra-se sobre as questões éticas, a necessidade de estabelecer um enquadramento jurídico, e uma especial atenção à área criminal face às experiências na esfera da justiça.

Deste modo, o presente artigo estrutura-se em três tópicos. O primeiro parte de uma revisão da literatura a partir da qual se destacam alguns dos principais riscos associados à aplicação da IA na esfera da justiça. O segundo tópico reúne os principais documentos produzidos pelas instâncias europeias como forma de responder aos desafios colocados, os quais colocam a tónica sobre a questão ética, a necessidade de regulação jurídica e uma atenção particular sobre a área criminal com base na justiça preditiva, face aos direitos fundamentais. No terceiro tópico, e quanto à questão da justiça preditiva, em particular, apresentam-se os exemplos dos programas COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) dos Estados Unidos, e do programa HART (Harm Assessment Risk Tool), do Reino Unido. Apresenta-se, ainda, de forma breve o contexto português e o posicionamento quanto à justiça preditiva com recurso à IA.

Os riscos da IA na justiça – uma (possível) revisão da literatura

Na política pública da administração da justiça, o debate sobre o uso das novas tecnologias e da IA, de um modo geral, gira em torno dos riscos e desafios que impõe na sua utilização. O consenso assenta sobre o reconhecimento de que algumas aplicações de IA, principalmente no tratamento de dados e informação, contribuem para uma justiça mais rápida e eficiente. As ferramentas de IA podem ser utilizadas no sistema de justiça com diversas finalidades: *i*) busca de jurisprudência avançada; *ii*) resolução de

disputas *online*; *iii*) análise preditiva de decisões; *iv*) triagem de processos; *v*) agrupamento por similaridade de jurisprudência; *vi*) transcrição de voz para textos com contexto; *vii*) geração semiautomática de peças; entre outras. Porém, o tratamento de decisões judiciais por IA, deve ser, também, encarado com reservas (Contini, 2020; Reiling, 2020). É sobre estas mesmas reservas que incidirá o presente tópico.

De acordo com a revisão da literatura quanto ao recurso da IA na administração da justiça, os problemas assentam sobre a forma como os algoritmos aplicados são selecionados e programados. 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*”) (Wisser, 2019). O processo de interpretação dos dados torna-se opaco, não havendo transparência no tratamento dos dados e como tal resulta na decisão, até mesmo para os próprios programadores dos algoritmos. No âmbito jurídico, é essencial que as novas tecnologias de auxílio à decisão judicial sejam capazes de funcionar como instrumento de apoio para um juiz humano, e nesse sentido, sejam capazes de fornecer a fundamentação exigida por lei. Porém, a opacidade envolvida no uso e construção dos sistemas inteligentes pode dificultar, ou mesmo tornar inviável, esse tipo de fundamentação (Maranhão et al., 2021).

Um outro risco que se associa à questão da transparência, são os “vieses algorítmicos” que dão origem a riscos de discriminação. O viés algoritmo ocorre quando as “máquinas refletem os valores humanos implícitos envolvidos na programação” (Nunes & Marques, 2018, p. 425). A IA estando dependente da intervenção humana, está submetida às opções éticas daqueles que a programam e alimentam o banco de dados que servirá de base para a tomada de decisões. Através do *machine learning*², os sistemas são capazes de se adaptarem a novas circunstâncias e de extrapolar padrões originais, fazendo com que certas dificuldades e divergências hermenêuticas acabem por ser reproduzidas num novo formato (Martins, 2021). Quer isto dizer, que se os dados são tendenciosos, as desigualdades estruturais e os preconceitos inculcados nos mesmos serão amplificados por meio da atividade da IA (cf. Silva & Ehrhardt Júnior, 2020), de que são exemplo o programa COMPAS e o programa HART, de que falaremos mais adiante.

O tratamento dos dados pela IA comporta, assim, o risco de as decisões poderem ser apresentadas como neutras e objetivas, quando na realidade não o são,

2 “*Machine learning*” (“aprendizagem da máquina”) é o processo de usar modelos matemáticos de dados para ajudar um computador a aprender sem instrução direta, permitindo que um sistema de computador continue a aprender e melhorar por conta própria, com base na experiência (IBM, s.d.).

colocando em risco a discricionariedade judicial, a independência e, em última análise, o julgamento justo, tornando-se uma questão institucional e constitucional (cf. Contini & Reiling, 2022). A interpretação do direito processual, na prática, é feita por quem cria, desenvolve, ou controla, o *software*, o qual com base em correlações estatísticas entre os dados, auxilia o juiz nas suas decisões. A questão é que este processo por se apresentar como “ciência dos dados”, pode ser falacioso quanto à influência que opera no juiz (cf. Contini & Reiling, 2022).

Existe, ainda, o risco sobre a segurança e a confiabilidade dos sistemas. A questão da responsabilidade pelos sistemas – a “*accountability*” – evidencia a importância do cumprimento das normas e da legislação para garantir o seu adequado funcionamento, nomeadamente, quanto à privacidade e proteção de dados pessoais, a propriedade dos algoritmos e, ainda, a segurança da informação e integridade de dados e sistemas (Ettikoven & Prins, 2018; Novelli et al., 2023). O uso indevido de dados pessoais e atentados à privacidade, destaca esta matéria devido ao facto da sua componente técnica estar diretamente ligada à ética (ex: discriminação), à confiança (ex: integridade dos sistemas) e à segurança (ex: *hacks* de dados e sistemas) (cf. Organisation for Economic Cooperation and Development [OECD], 2023).

O contexto europeu – Conselho da Europa e União Europeia

No contexto europeu, face ao exposto, as principais preocupações recaem sobre a definição de um quadro jurídico regulatório da conceção, desenho e aplicação da IA de modo a garantir o respeito pelos direitos fundamentais e os princípios de um Estado de Direito democrático. As diferentes instâncias europeias, como o Conselho da Europa, a União Europeia (UE) e suas instituições, têm produzido diversos documentos e orientações nesse sentido, alguns dos quais se destacam na Tabela 1, aqui apresentada.

Quanto aos direitos fundamentais, vale a pena assinalar a carta ética da Comissão Europeia para a Eficiência da Justiça, a CEPEJ, a qual teve o mérito de elaborar o primeiro documento sobre esta temática – *Carta Ética sobre o Uso de Inteligência Artificial em Sistemas Judiciais* (Comissão Europeia para a Eficiência da Justiça [CEPEJ] 2018) –, demonstrando, também, a sua importância quanto à área da justiça e identificando cinco princípios éticos: *i*) Princípio de respeito aos direitos fundamentais; *ii*) Princípio de não-discriminação; *iii*) Princípio de qualidade e segurança; *iv*) Princípio da transparência, imparcialidade e equidade; e *v*) Princípio “sobre o controlo do utilizador”. Estes procuram dar resposta aos principais desafios colocados pela utilização da IA na justiça, os quais foram apurados pelas

Tabela 1 Documentos produzidos pelas instâncias europeias

Conselho da Europa	União Europeia
1) Carta Ética sobre o Uso de Inteligência Artificial em Sistemas Judiciais (2018) pela Comissão Europeia para a Eficiência da Justiça (CEPEJ)	1) Criação do Grupo de peritos de alto nível sobre a IA (GPAN IA)
2) Possível introdução de um mecanismo de certificação de ferramentas e serviços de inteligência artificial na esfera da justiça e do judiciário: Estudo de Viabilidade (2020)	2) Orientações éticas PARA UMA IA DE CONFIANÇA (2019) (pelo GPAN— IA)
3) Roteiro revisto para garantir um acompanhamento adequado da Carta Ética da CEPEJ sobre o uso de inteligência artificial nos sistemas judiciais e seu ambiente (2021)	3) LIVRO BRANCO sobre a inteligência artificial— Uma abordagem europeia virada para a excelência e a confiança (2020)
4) CAHAI— Comitê ad-hoc de Inteligência (2019-2021) Possíveis elementos de um quadro jurídico sobre inteligência artificial baseado nas normas do Conselho da Europa sobre direitos humanos, democracia e Estado de direito (2021)	4) Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (regulamento inteligência artificial) e altera determinados atos legislativos da União (2021) — "AI ACT" (última versão aprovada a 13 de março de 2024)
5) CAI— Comitê de Inteligência Artificial (CAI) (2022-2024)	5) Acesso à Justiça— Aproveitando as Oportunidades da Digitalização" (2020)
	6) Proposta de Diretiva de Responsabilidade por Inteligência Artificial (AILD) (2022)

Fonte: Autores.

experiências dos programas aplicados, seja pela análise de *big data* como, principalmente, por alguns programas baseados em justiça preditiva.

Paralelamente, a UE, através da criação do grupo de peritos de alto nível, seguiu um raciocínio semelhante com a elaboração das *Orientações éticas para uma IA de confiança* (Grupo de Peritos de Alto Nível Sobre a Inteligência Artificial [GPAN IA], 2019). Embora com um carácter mais geral, no relatório são identificados os sete principais requisitos: "1) Ação e supervisão humanas; 2) Solidez técnica e segurança; 3) Privacidade e governação dos dados; 4) Transparência; 5) Diversidade, não discriminação e equidade; 6) Bem-estar societal e ambiental; e 7) Responsabilização." (GPAN IA, 2019, pp. 17-18). Estes requisitos mais não são do que uma forma de respeitar os direitos e liberdades fundamentais, pedra angular para ambas as instâncias europeias. A conceção, programação e aplicação da IA deve estar sujeita à supervisão dos seres humanos, e ao ser concebida e aplicada por estes, deve igualmente seguir padrões éticos que sejam universalmente reconhecidos. O importante será sempre o respeito e a segurança dos seres humanos.

De mencionar também a preocupação com a regulação política e jurídica da IA. O Artificial Intelligence Act— AI ACT (2024, p. 1) tem como:

objetivo a melhoria do funcionamento do mercado interno mediante a previsão de um regime jurídico uniforme, (...) em conformidade com os valores da União, a fim de promover a adoção de uma inteligência artificial (IA) centrada no ser humano e de confiança.

O regulamento foi recentemente aprovado pelo Parlamento Europeu, em março de 2024, marcando um dos derradeiros passos de uma iniciativa que teve início em 2021. É de realçar que o AI ACT é a primeira iniciativa regulatória global e abrangente, com o objetivo de se constituir, após o seu processo de aprovação, num instrumento de *hard law* vinculativo para todas as pessoas e entidades públicas e privadas produtoras ou usuárias de IA no espaço dos 27 Estados da UE.

Assente numa abordagem baseada no risco (risco inaceitável, alto risco, risco limitado e risco mínimo), os sistemas de IA de “alto risco” são os que criam riscos significativos para a saúde e a segurança ou para os direitos fundamentais das pessoas, os quais terão de cumprir um conjunto de requisitos obrigatórios horizontais para uma IA de confiança e seguir procedimentos de avaliação da conformidade antes de poderem ser colocados no mercado da União (AI ACT, 2024). Incluem-se aqui os sistemas de IA:

concebidos para a administração da justiça e os processos democráticos (...) tendo em conta o seu impacto potencialmente significativo na democracia, no Estado de direito e nas liberdades individuais, bem como no direito à ação e a um tribunal imparcial. Em particular, para fazer face aos riscos de potenciais enviesamentos, erros e opacidade, é apropriado classificar como sendo de risco elevado os sistemas de IA concebidos para serem utilizados por uma autoridade judiciária ou para, em seu nome, auxiliar autoridades judiciárias na investigação e interpretação de factos e do direito e na aplicação da lei a um conjunto específico de factos. (AI ACT, 2024, p. 61)

Entre as práticas proibidas no espaço europeu encontram-se os sistemas de categorização biométricos que são baseados em características sensíveis ou, ainda, a proibição do uso de imagens de rostos retiradas da *internet* ou de sistemas de videovigilância para criar bases de dados de reconhecimento facial; o uso de sistemas de reconhecimento de emoções no local de trabalho ou escolas, a criação de *rankings* sociais, ou o que se considera “policimento preditivo”, quando é baseado apenas no perfil das pessoa ou na avaliação das suas características; e o uso de IA que possa manipular o comportamento humano ou explorar as vulnerabilidades dos utilizadores (cf. AI ACT, 2024).

No que à parte criminal diz respeito, evidencia-se que o exercício de importantes direitos fundamentais processuais, como o direito à ação e a um tribunal imparcial, o direito à defesa e a presunção de inocência, não pode ser prejudicado,

em particular, pela falta de transparência e explicabilidade dos sistemas de IA. Especificamente quanto aos sistemas de IA de alto risco concebidos para serem utilizados por autoridades de aplicação da lei, ou em seu nome, ou por órgãos ou organismos da União em apoio das autoridades de aplicação da lei, são realçados os que servem:

para avaliar o risco de uma pessoa singular vir a ser vítima de infrações penais, como polígrafos e instrumentos semelhantes, para avaliar a fiabilidade dos elementos de prova no decurso da investigação ou da repressão de infrações penais, e, na medida em que tal não seja proibido nos termos do presente regulamento, para avaliar o risco de uma pessoa singular cometer uma infração ou reincidência não apenas com base na definição de perfis de pessoas singulares ou na avaliação de os traços e características da personalidade ou do comportamento criminal passado de pessoas singulares ou grupos, para a definição de perfis no decurso da deteção, investigação ou repressão de infrações penais. (AI ACT, 2024, p. 58)

É notória a preocupação quanto ao impacto da utilização da IA na justiça criminal, por ser nesta área cuja ética, controlo e segurança, elementos transversais em todos os documentos produzidos, mais se revelam problemáticas na aplicação da lei e harmonização da jurisprudência. Existe uma preocupação particular com a área penal, realçando-se que alguns campos do direito são mais sensíveis do que outros, devido às ramificações das decisões judiciais para as liberdades pessoais, pelo que a criminologia e o direito penal são geralmente identificados como áreas do direito onde o uso da IA deve ser encarado com a máxima cautela (Comissão Europeia para a Eficiência da Justiça [CEPEJ], 2020).

É importante referir que na já mencionada *Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente* (CEPEJ, 2018), adotada pela CEPEJ, pode ler-se na introdução que o

tratamento de decisões judiciais por inteligência artificial, de acordo com os seus criadores, é suscetível, em matéria civil, comercial e administrativa (...) Em matéria penal, a sua utilização deve ser considerada com as maiores reservas, a fim de evitar a discriminação com base em dados sensíveis, em conformidade com as garantias de um julgamento justo. (CEPEJ, 2018, p. 4)

Relativamente à matéria criminal, a avaliação é a de que:

mesmo não sendo especificamente concebidas para serem discriminatórias, a utilização de estatísticas e de IA nos processos penais mostrou um risco de provocar o ressurgimento de doutrinas deterministas em detrimento de doutrinas de individualização da sanção, que foram amplamente adquiridas desde 1945 na maioria dos sistemas judiciais europeus. (CEPEJ, 2018, p. 39)

Nesta senda, o Comité Europeu sobre problemas de crime (European Committee on Crime Problems – CDPC)³, o qual coordena as atividades do Conselho da Europa na área da prevenção e controlo do crime, tem depositado sobre a IA uma atenção redobrada. No estudo *Feasibility Study on a future Council of Europe instrument on artificial intelligence and criminal law*, e devido à implementação dos sistemas de IA e a ocorrência dos primeiros incidentes e acidentes, ficou claro que na área da justiça criminal, não apenas a alocação da responsabilidade, como também o respeito por certos direitos fundamentais, em particular, os direitos judiciais, constituem-se numa questão peculiar (European Committee on Crime Problems [CDPC], 2020). Em causa estão, principalmente, o respeito pelo direito a um julgamento justo, consagrado nos artigos 5.º, 6.º e 7.º da Convenção Europeia dos Direitos Humanos (CEDH), juntamente com o respeito pela vida privada, salvaguardado pelo artigo 8.º da CEDH, dado que são afetados quando os dados coletados e usados pelos sistemas de IA podem ser oferecidos como prova em processos criminais.

Em outubro de 2021, no que diz respeito à UE, o Parlamento Europeu aprovou a resolução sobre *A inteligência artificial no direito criminal e a sua utilização pelas autoridades policiais e judiciárias em questões criminais*. O documento reforça que a utilização de aplicações de IA no domínio da justiça e da aplicação da lei:

sejam tomadas de forma transparente, salvaguardando na íntegra os direitos fundamentais e, em particular, não perpetuem a discriminação, a parcialidade ou os preconceitos, onde quer que existam; considerando que as opções políticas pertinentes devem respeitar os princípios da necessidade e da proporcionalidade, de modo a garantir a constitucionalidade e um sistema judicial equitativo e humano. (Parlamento Europeu [PE], 2021b, p. 4)

Pela parte da UE é, ainda, pertinente mencionar o relatório conjunto da eu-LISA – a agência da UE para a gestão operacional de sistemas informáticos de grande escala na área de liberdade, segurança e justiça – e do EUROJUST – a agência da UE para a cooperação da justiça criminal –, intitulado *Artificial Intelligence – supporting cross-border cooperation in criminal justice* (eu-LISA & EUROJUST, 2022). O relatório assinala os seus efeitos positivos, identificando situações em que IA pode ser aplicada efetivamente, resultando em ganhos significativos sem implicações éticas ou legais de grande relevância. Contudo, também concluem que o desenvolvimento e a implantação de soluções baseadas em IA devem ser avaliados com cuidado especial, a fim de evitar situações em que os direitos e liberdades fundamentais dos indivíduos sejam afetados negativamente e as pessoas sejam submetidas a várias

3 Cf. European Committee on Crime Problems (CDPC, s.d.).

formas de discriminação, limitações de direitos, tratamento injusto, etc. Para tal, é sugerida uma abordagem baseada no risco, integrando uma análise de custos-benefícios, direitos fundamentais e avaliação de proteção de dados, por forma a determinar aspetos a salvaguardar na aplicação e implementação dos sistemas de IA na prática (cf. eu-LISA & EUROJUST, 2022).

Justiça preditiva: o exemplo do COMPAS e do HART

A predição consiste no uso de dados anteriores, algoritmos estatísticos e técnicas de *machine learning* para identificar a probabilidade de resultados futuros, o que coloca questões éticas relativamente à previsão das decisões (Godefroy et al., 2019; Reiling, 2020). Pode afirmar-se que a IA dá origem a uma forma de “justiça preditiva”, através da “definição de perfis pessoais discriminadores”. O uso da IA em programas na área criminal, com a utilização de algoritmos para traçar o perfil dos indivíduos alegadamente delinquentes pode representar 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". (CEPEJ, 2018, p. 39)

Vejamos então, dois exemplos desta justiça preditiva.

O primeiro, o COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) nos Estados Unidos da América (EUA), concebido pela empresa Northpointe e utilizado na justiça americana, é um programa de cálculo do risco de reincidência criminal e de auxílio na decisão judicial.

Este inclui 137 perguntas⁴ respondidas pelo réu, ou informações extraídas de registos criminais, nomeadamente, se tem dificuldade para pagar contas, histórico familiar, histórico criminal do acusado, etc., e em que o algoritmo classifica a pessoa numa escala de 1 (baixo risco) a 10 (alto risco), sendo que as suas conclusões auxiliam o juiz a decidir sobre a sentença. O algoritmo preocupa-se principalmente com o risco de reincidência, destinando-se a ser utilizado na tomada de decisões de libertação pré-julgamento (fiança) e liberdade condicional. Uma variável que não aparece é a raça do réu, contudo, muitas das variáveis utilizadas no COMPAS estão fortemente correlacionadas com a raça, a qual está incorporada

4 Cf. ProPublica (s.d.).

na estrutura e na história do sistema de justiça criminal nos Estados Unidos (Fraenkel, 2020).

De acordo com um estudo levado a cabo pela ONG ProPublica, o programa erroneamente coloca sob o risco de incidência criminal uma maior taxa de pessoas de origem afro-americana, o qual foi identificado como duas vezes superior relativamente a outras populações (cf. CEPEJ, 2018). Podem indicar-se duas principais fontes que contribuem para discriminação racial: o próprio preconceito inerente do programador que é inculcado na criação do algoritmo; e os diversos fatores que o algoritmo leva em consideração sobre os réus (Zhang & Han, 2022). Quanto ao primeiro, muitas vezes o programador não se apercebe que existe preconceito, tornando o algoritmo tendencioso. Da mesma forma, a raça não deveria ser levada em consideração ao projetar os algoritmos, no entanto, se os programadores incluírem a raça como critério nos algoritmos, diferentes raças receberão pontuações diferentes, resultando em preconceito racial. Além disso, se a maioria, ou quase todos, os programadores forem da mesma raça, as questões raciais perdem relevância (cf. Zhang & Han, 2022). No que diz respeito à segunda fonte, trata-se de fatores estáticos, isto é, os fatores que não podem ser alterados, como a raça, o histórico profissional, a educação e o histórico criminal dos réus. Estes referem-se à história dos indivíduos, portanto, não podem mudar, e uma vez que não mudarão no futuro, utilizar a história passada dos arguidos para prever as suas ações pode conter preconceito. Por exemplo, os indivíduos com ensino superior receberiam pontuações de risco relativamente mais baixas no COMPAS do que os indivíduos sem escolaridade; os réus com trabalho a tempo inteiro e salários elevados receberiam pontuações de risco mais baixas do que aqueles sem trabalho. Deste modo, muitos arguidos provêm de um contexto de vulnerabilidade socioeconómica, o que resulta numa pontuação de risco mais elevada do que os arguidos mais privilegiados (cf. Zhang & Han, 2022).

Vários críticos levantaram a possibilidade de que os modelos de computador que aprendem padrões de dados podem ser sutilmente tendenciosos contra certos grupos com base em preconceitos embutidos nesses dados (Surden, 2019). Nesta perspetiva, outros autores afirmam que a abordagem racional, científica e “neutra em termos de valor” para a condenação funciona para obscurecer as formas pelas quais a avaliação de risco atribui um estigma a certas características individuais. A questão é a de que a tomada de decisões tendenciosa, ou preconceituosa, é meramente substituída por um procedimento padronizado que tem um preconceito inerente contra a marginalidade socioeconómica. Em resumo, a racionalização funciona para legitimar a estigmatização e, portanto, a reprodução da marginalização e da desigualdade através de procedimentos de condenação baseados no risco (cf. van Eijk, 2017).

Tal só veio demonstrar a “fragilidade social e económica de certos grupos de populações (...) [e] a falta de transparência nos processos de operação dos algoritmos concebidos por empresas privadas” (CEPEJ, 2018, p. 41)⁵.

Numa análise do programa COMPAS, Contini (2020) assinala que se trata de uma autoridade extrajudicial (uma empresa privada) que inscreveu no sistema as regras a serem adotadas para calcular os riscos de reincidência. Tais regras são fechadas, “*black boxed*”, justificadas em termos gerais e inacessíveis aos juizes, ou aos envolvidos no processo. Uma vez instalado o sistema, é provável que o utilizador delegue à máquina a decisão sobre a prisão preventiva. A questão que se impõe é: será que o juiz está disposto a ir contra um cálculo de avaliação de risco feito por uma máquina? (Contini, 2020). Num cenário diferente, o juiz pode decidir manter um réu sob custódia devido à sua avaliação do risco de reincidência, mas o sistema de IA pode prever uma baixa probabilidade de reincidência. Estará o juiz disposto a ignorar a sugestão da máquina e manter o arguido preso? O risco de que o sistema aumente o número de falso-positivos é alto; isto é, pessoas mantidas sob custódia que poderiam ter sido libertadas. O uso de tais sistemas fornece uma influência nova e indevida na tomada de decisões judiciais, colocando em risco a independência judicial e o julgamento justo. As “sugestões” de sistemas inescrutáveis tornam-se influências indevidas na tomada de decisões judiciais, afetando a independência e a justiça judicial (cf. Contini, 2020).

O segundo, o programa HART (Harm Assessment Risk Tool), desenvolvido em parceria com especialistas em estatística da Universidade de Cambridge e a polícia de Durham, enquanto programa de auxílio quanto à decisão policial na avaliação de detidos sobre a reincidência. O objetivo era permitir que os presos classificados como de risco moderado — com a possibilidade de reincidência de crime, mas sem ser violento — fossem elegíveis para o programa Checkpoint — uma “disposição extrajudicial”, isto é, uma forma de lidar com uma ofensa que não requer processo em tribunal, destinada a reduzir futuros delitos, adaptando estratégias de intervenção que possam ajudar os indivíduos a deixar a “vida do crime”.

O programa HART, que está a ser testado desde 2017, tomou como base a aprendizagem automática usando arquivos da Polícia de Durham de 2008 a 2012 (104.000 casos), cuja aprendizagem tomou como base as decisões policiais, e a partir desta informação, foi classificado o risco de reincidência (baixo, moderado ou

5 Deve referir-se que as conclusões da ProPublica foram criticadas, tanto por motivos técnicos relacionados com a validade da análise estatística como em relação à forma como as conclusões foram relatadas. No entanto, mesmo um autor de uma destas refutações críticas sugeriu que “o que parece ser tendencioso não está na ferramenta — está no sistema”. Tal parece reintroduzir a noção de “*techwashing*” — e as consequências prejudiciais e discriminatórias para os indivíduos são praticamente as mesmas, independentemente da descrição do problema (cf. Cilevičs, 2020, p. 12).

alto), sem que os dados tivessem necessariamente relacionados com a natureza do crime em causa. A maior parte desses fatores concentram-se no histórico de comportamento criminoso do indivíduo, ao que acresce o número de relatórios policiais sobre o indivíduo, e outras variáveis como a idade, o sexo e dois tipos de código postal residencial. Este último fator é relevante, dado que uma investigação levada a cabo pelo Big Brother Watch (2020), uma organização de defesa de liberdades civis e de direitos humanos, identificou que uma das variáveis de código postal inseridas no sistema HART resultava de um produto de dados de marketing comercial da empresa de dados global Experian, conhecido como “*Mosaic*”. O “*Mosaic*” é uma ferramenta de segmentação social, geográfica e demográfica, que consiste em estereótipos de códigos postais criados a partir de 850 milhões de dados, incluindo dados dos censos, etnia, dados de saúde, emprego, benefícios infantis e subsídios de apoio ao rendimento, nomes familiares e pessoais vinculados à etnia, dados extraídos de fontes *online*, entre outros (cf. Big Brother Watch, 2020). Estes dados são utilizados para traçar o perfil de 50 milhões de adultos no Reino Unido, com base nos seus códigos postais, criando perfis estereotipados através da atribuição de características demográficas, os quais, em 2018, incluíam categorias como “Herança Asiática”, “Juventude Desconectada”, “Caleidoscópico Lotado”, “Famílias com Necessidades” ou “Trabalhadores com baixo salário”. A título ilustrativo, o “Caleidoscópico Lotado” foi descrito como famílias “multiculturais” que provavelmente vivem em apartamentos “apertados” e “sobrelotados”, com nomes como “Abdi” e “Asha”, e as “Famílias com Necessidades” foram descritas como recebendo “uma série de benefícios” com nomes como “Stacey”, enquanto os “Trabalhadores com baixo salário” foram tipificados como tendo “poucas qualificações” e “assíduos espetadores de televisão”, com nomes como “Terrence” e “Denise” (cf. Big Brother Watch, 2020). A investigação realçou que permitir este tipo de dados de perfil, que inclui não apenas dados étnicos, mas toda uma série de outras informações socioeconómicas e raciais, sejam utilizados em algoritmos do sector público é discriminatório e, no sistema de justiça criminal, conduziria a práticas injustas e decisões imprecisas⁶. Os métodos utilizados pelo programa HART foram alvo de críticas, por constituir uma abordagem estatística de uma amostra de dados, que conduz a um resultado de discriminação negativa de certos indivíduos e grupos (Reiling, 2020).

Na ótica da investigação criminal, as decisões com base nos sistemas de IA colidem com diferentes liberdades fundamentais, nomeadamente, a presunção de inocência. A nova linguagem da matemática, que ajuda a definir novas categorias

6 Após a investigação do sistema HART pela Big Brother Watch, a Polícia de Durham removeu os dados do Experian Mosaic ainda em 2018.

como “pessoa de interesse”, redireciona as atividades das instituições que aplicam o direito para indivíduos ainda não considerados “suspeitos”. As novas noções que estão a ser criadas contrariam os padrões de prova estabelecidos no processo penal (Završnik, 2020). Segundo Završnik (2020, p. 575), os seus efeitos podem ser observados relativamente a princípios como a igualdade e discriminação, nos quais o “*over-policing*” (“sobre-policamento”; “policamento excessivo”), derivado de programas de “justiça preditiva”, ocorre quando a polícia patrulha áreas com maior taxa de crimes, o que por sua vez amplia a necessidade de policamento de áreas já policiadas. Este é um excelente exemplo do efeito de “círculo vicioso” do uso do *machine learning* no domínio do controlo do crime. Contudo, o “sob-policamento”, isto é, quando a polícia não escrutina certas áreas na mesma medida que outras, torna-se ainda mais problemático, conduzindo a um sentimento e experiência não proporcional de justiça. Alguns tipos de crime são então mais propensos a serem julgados do que outros e o princípio central da legalidade — a exigência de que todos os crimes sejam julgados *ex officio*, por oposição ao princípio da oportunidade, não é respeitado.

Os criminologistas críticos mostraram como a própria definição de crime — e ainda mais a acusação do crime — é inerentemente política: as instituições que aplicam o direito/lei são forçadas a tomar decisões “políticas” sobre qual crime processar e investigar, devido aos recursos materiais e humanos serem limitados. Estes priorizam as atividades de forma explícita ou implícita. A desigualdade no policamento preditivo altera, portanto, a perceção do que é considerado como “crime grave” em primeiro lugar (cf. Završnik, 2020). Em suma, analisam-se dados históricos relativos a taxas de reincidência de amostras de criminosos, o que permite determinar quais os fatores que estão estatisticamente relacionados com a reincidência, com o escrutínio de dados pessoais que incluem a sua vivência familiar, atividade profissional, estabilidade familiar, entre outros fatores preditivos como o género, a raça e a classe social, os quais são incorporados num algoritmo estatístico que pesa mais fortemente uns fatores em relação a outros (cf. Alberto, 2024). O seu impacto pode, portanto, ser considerado paradoxal quando aplicado no processo penal, dado que, se por um lado, a utilização da IA pode ser útil para prevenir e avaliar o risco criminal, por outro lado, estes avanços tecnológicos na justiça podem ser considerados invasivos e:

potencialmente suscetíveis de afetar esferas de direitos de indivíduos, fronteiras intransponíveis para o Estado, e que poderão restringir tragicamente a função da privacidade, não apenas ao nível do livre desenvolvimento da personalidade individual, como também ao nível de manutenção de uma democracia viva. (Alberto, 2024, p. 8)

Assim, na área penal, pode afirmar-se:

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, a tecnologia pode acabar por legitimá-las. (CEPEJ, 2018, p. 43)

O contexto português quanto à justiça preditiva

Na Europa continental pode afirmar-se que ainda é rara a utilização de instrumentos preditivos. Sobre esta perspetiva, Anabela Rodrigues (2023) afirma que tal se deve ao modelo de justiça penal dos sistemas anglo-saxónicos com reflexo nas finalidades da punição, e onde continua a ter destaque a fase de julgamento (de discussão da questão da culpa), na qual a intervenção humana é essencial. Para além disso, acrescenta que no âmbito do:

Regulamento Geral da Proteção de Dados (RGPD) e o Regulamento que estabeleceu a Autoridade Europeia para a Proteção de Dados (...), no ordenamento jurídico português, que a Lei n.º 58/2019, de 8 de agosto, que assegura a execução do RGPD, é mais protetora das pessoas do que o RGPD, uma vez que *não prevê* que o consentimento do titular dos dados legitime a sua sujeição a uma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis. (Rodrigues, 2023, pp. 228-229)

Quer isto dizer que é proibido o recurso exclusivo ao tratamento de dados pessoais para fundamentar decisões, bem como outras finalidades no âmbito do processo penal (cf. Rodrigues, 2023).

Em Portugal, a área da justiça tem vindo a reforçar o esforço de modernização, simplificação e transformação digital, de que é exemplo a Estratégia GovTech da Justiça, “materializa-se num conjunto de projetos de inovação e transformação digital, desenvolvidos em colaboração com universidades, centros de investigação e startups com o objetivo de tornar os serviços da justiça mais ágeis e eficientes” (PORTUGAL.GOV.PT, 2023, para. 1). O recurso à IA no sistema de justiça português, toma como base um modelo de “*machine learning*”, com o objetivo de melhorar a informação prestada quanto aos serviços disponibilizados na *internet* (Lusa, 2023a). A nova tecnologia será utilizada para validar a autenticidade dos documentos necessários aos pedidos efetuados, nomeadamente, em processos de concessão de nacionalidade, sendo também aplicada na anonimização de decisões judiciais procedendo automaticamente à remoção de dados pessoais e aliviando,

assim, a carga de trabalho manual dos funcionários judiciais (Lusa, 2023a). Encontra-se, igualmente, em desenvolvimento, uma ferramenta de IA para ajudar no aconselhamento jurídico denominada *Guia prático da Justiça* (Secretaria-Geral do Ministério da Justiça [SGM], s.d.), que consiste numa espécie de “Chat GPT” (com uma componente de IA com base no modelo de linguagem, e usando técnicas de *deep learning* para conseguir criar uma linguagem natural) para questões simples sobre temas como o casamento, divórcio e criação de empresas.

Os programas de apoio à decisão judicial ainda não são uma realidade em Portugal. Porém, e mesmo que não existam ainda, no que diz respeito ao recurso à IA no apoio às decisões judiciais, a perspetiva do poder judicial é a de que a utilização das tecnologias de IA não pode substituir o papel do juiz na formação da decisão, e não deve poder enviesar ou interferir negativamente na fundamentação das decisões (Lameira, 2021). No âmbito de uma “reforma da justiça”, e contemplando inevitavelmente o impacto da IA, existe o consenso de que a sua aplicação pode servir como uma espécie de assistente judicial eletrónico, para, por exemplo, preparar e auxiliar o juiz na redação das decisões finais, essencialmente, servindo para reduzir o trabalho material do julgador em várias matérias, como na indexação e classificação dos dados constantes do processo eletrónico ou dos seus anexos, no cálculo de indemnizações, etc. (cf. Coelho et al., 2023). Contudo, quanto à análise jurisprudencial preditiva, e quanto a processos mais complexos e em que estejam em causa interesses mais relevantes socialmente como, por exemplo, um julgamento criminal, a perspetiva é a de que a IA não faz sentido, reforçando que a justiça deve ser feita pelos juízes (cf. Coelho et al., 2023). A visão é a de que influência da IA na tomada da decisão final deve ser clara e transparente, no sentido que possa ser posta em causa pelas partes no processo e contestada em recurso judicial, no âmbito do Estado de Direito e dos direitos fundamentais, garantindo uma adequada articulação entre tecnologia, direitos humanos e direitos fundamentais.

No entanto, na área criminal, pode mencionar-se um projeto piloto, criado por uma empresa portuguesa (a IPS – Inovative Prison Systems), referente à gestão penitenciária, com recurso à IA e à análise preditiva, com o objetivo de ajudar na tomada de decisões na área da justiça criminal, nomeadamente, o *software* de administração “Horus 360° iOMS” (Intelligent Offender Management System)⁷. Trata-se de um *software* de administração que integra IA em todo o processo na justiça, fornecendo recomendações de medidas a aplicar com base no perfil da pessoa. O projeto assume-se como um “Sistema Inteligente de Gerenciamento de Infratores” que permite gerenciar o ciclo de vida dos cidadãos em custódia ou supervisionados no

7 Cf. Revista JUSTICE TRENDS (s.d.).

âmbito de processos criminais (prisão e liberdade condicional), e apoiar a tomada de decisões no âmbito de detenção, ressocialização, intervenção terapêutica, decisões judiciais, decisões sobre medidas de segurança, tratamento ou libertação antecipada da prisão. O elemento inovador é o recurso à IA ao longo de todo o processo, desde que o indivíduo em conflito com a lei é detido, passando pelo julgamento, até ao estabelecimento prisional e às recomendações de programas de reabilitação, dependendo do perfil e da disponibilidade dos programas que possam existir (Lusa, 2023b). Embora não esteja a ser testado em Portugal, encontra-se em “fase piloto” em diferentes países europeus, com interesse para várias entidades que operam na área da justiça criminal, desde a polícia, ao tribunal ou às prisões, contudo, requerendo sempre uma adaptação às jurisdições específicas (Lusa, 2023b).

Conclusão

A chegada da IA à administração da justiça trouxe consigo novas oportunidades e desafios. Pela sua dimensão disruptiva, os maiores riscos relacionam-se com a ética, a transparência e a segurança na utilização dos dados. A área da justiça é particularmente sensível a estas questões devido ao seu papel de garante dos direitos fundamentais e dos princípios do estado de direito, com destaque para a área criminal, na qual a justiça preditiva com recurso à IA pode conduzir, entre outros, à violação do direito de igualdade perante a lei, da dignidade humana, da não discriminação, da proteção de dados pessoais e da vida privada, e do direito à ação e a um tribunal imparcial.

Nesse sentido, a Europa e suas instâncias têm procurado salvaguardar através da produção de diversas recomendações e orientações (*soft law*), que:

todas as soluções policiais e judiciais baseadas na inteligência artificial devem também ser utilizadas no pleno respeito pela dignidade humana, pelos princípios da não discriminação, da liberdade de circulação, da presunção de inocência e do direito de defesa, incluindo o direito ao silêncio, a liberdade de expressão e o livre acesso à informação, a liberdade de reunião e a liberdade de associação, a igualdade perante a lei, o princípio da igualdade das partes e o direito a um recurso efetivo e a um julgamento justo (...) deve ser proibida toda e qualquer utilização de IA que seja incompatível com os direitos fundamentais. (Parlamento Europeu [PE], 2021a, p. 8)

Das experiências analisadas neste artigo (COMPAS e HART), fica claro que a inclusão de algoritmos nas decisões judiciais criminais através da justiça preditiva pode alterar o próprio carácter dos dados, enviesando de forma discriminatória os resultados. Da mesma forma, sendo impregnados de uma suposta objetividade e imparcialidade,

podem condicionar a independência dos juízes no seu processo decisório e, por conseguinte, o direito a um julgamento justo.

Ora, se na Europa, a maioria dos projetos de IA para utilização na administração da justiça criminal são, ainda, pilotos, em Portugal, estamos, ainda, perante uma realidade distante. Porém, é inegável o ritmo com que estas novas tecnologias se desenvolvem, e a inevitabilidade da sua expansão e uso nas diferentes esferas da justiça, incluindo a criminal. Assim, a concluir, identificamos três pontos essenciais numa agenda futura de reflexão e de análise sobre a utilização da IA na justiça criminal e a sua conformidade (ou não) com os direitos fundamentais: *i*) os processos de decisão e de ação dos decisores públicos e dos atores judiciais na defesa da produção e uso destas tecnologias, ao nível da governação política global, europeia e em cada Estado; *ii*) o recurso à interdisciplinaridade como epistemologia fundamental para a compreensão da aplicação do direito, o que convoca outras questões como o futuro da educação e formação jurídica, e da relação entre ciência e direito; *iii*) e, por último, as tecnologias de IA de natureza preditiva que venham a ser utilizadas na justiça criminal, mesmo no respeito dos limites éticos e das proibições jurídicas, quando vigentes, nomeadamente, na UE, que impactos e transformações irão produzir na racionalidade, natureza e modo de administrar justiça.

Nota

Por decisão pessoal, os/as autores/as do texto escrevem segundo o novo acordo ortográfico.

Referências

- AI ACT, Artificial Intelligence Act. (2024). *Regulamento Inteligência Artificial – Resolução legislativa do Parlamento Europeu, de 13 de março de 2024, sobre a proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da união (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD) – P9_TA(2024)0138*. Parlamento Europeu.
- Alberto, N. (2024, 20 de fevereiro). Inteligência Artificial e Decisão Judicial. *Observatório Almedina*.
<https://observatorio.almedina.net/index.php/2024/02/20/inteligencia-artificial-e-decisao-judicial/>
- Biard, A., Hoevenaars, J., Kramer, X., & Themeli, E. (2021). Introduction: The Future of Access to Justice – Beyond Science Fiction. In X. Kramer, A. Biard, J. Hoevenaars, & E. Themeli (Eds.), *New Pathways to Civil Justice in Europe* (pp. 1-20). Springer.

- Big Brother Watch. (2020). *Big Brother Watch briefing on Algorithmic Decision-Making in the Criminal Justice System*.
- Brito, T., & Fernandes, R. (2020). Inteligência Artificial e a Crise do Poder Judiciário: Linhas Introdutórias sobre a Experiência Norte-Americana, Brasileira e sua Aplicação no Direito Brasileiro. *Revista Acadêmica da Faculdade de Direito do Recife*, 91(2), 84-107.
- Campbell, R. W. (2020). Artificial intelligence in the courtroom: the delivery of justice in the age of machine learning. *Colo. tech. L.J.*, 18(2), 323-350.
- CDPC, European Committee on Crime Problems. (s.d.) *Home*. Consultado de <https://www.coe.int/en/web/cdpc>
- CDPC, European Committee on Crime Problems. (2020). *Feasibility Study on a future Council of Europe instrument on artificial intelligence and criminal law*. Council of Europe.
- CEPEJ, Comissão Europeia para a Eficiência da Justiça. (2018). *Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente*. Council of Europe.
- CEPEJ, Comissão Europeia para a Eficiência da Justiça. (2020). *Possible introduction of a mechanism for certifying artificial intelligence tools and services in the sphere of justice and the judiciary: Feasibility Study*. Council of Europe.
- Cilevičs, B. (2020). *Justice by algorithm – the role of artificial intelligence in policing and criminal justice systems* [Report – Committee on Legal Affairs and Human Rights]. Council of Europe.
- Coelho, N. (Coord.), Lopes, J. M., Coelho, A. de A., Martins, J. J. O., Latas, A. J., & Moura, S. (2023). *Agenda da Reforma da Justiça – Uma reflexão aberta e alargada do judiciário*. Almedina.
- Contini, F. (2020). Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings. *Law, Technology and Humans*, 2(1), 4-18.
- Contini, F., & Reiling, D. (2022). Double normalization: When procedural law is made digital. *Oñati Socio-Legal Series*, 12(3), 654-688.
- Donoghue, J. (2017). The Rise of Digital Justice: Courtroom Technology, Public Participation and Access to Justice. *The Modern Law Review*, 80(6), 995-1025.
- Ettekoven, B., & Prins, C. (2018). Data analysis, artificial intelligence and the judiciary system. In V. Mak, E. Tjong Tjin Tai, & A. Berlee (Eds.), *Research handbook in data science and law* (pp. 425-447). Edward Elgar.
- eu-LISA, European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice., & EUROJUST, The European Union Agency for Criminal Justice Cooperation. (2022). *Artificial Intelligence – supporting cross-border cooperation in criminal justice – Joint Report Prepared By eu-LISA and EUROJUST*. Publications Office of the European Union.
- Fabri, M., & Contini, F. (Eds.) (2001). *Justice and Technology in Europe: How ICT is changing the judicial business*. Kluwer Law International

- Fraenkel, A. (2020). *COMPAS Recidivism Algorithm* [Lecture Notes for UCSD course DSC 167 – Fairness and Algorithmic Decision Making].
<https://afraenkel.github.io/fairness-book/content/04-compas.html>
- Godefroy, L., Lebaron, F., & Lévy-Vehel, J. (2019). *Comment le numérique transforme le droit et la justice vers de nouveaux usages et un bouleversement de la prise de décision – Anticiper les évolutions pour les accompagner et les maîtriser*. GIP – Mission de recherche Droit et Justice.
- GPAN IA, Grupo de Peritos de Alto Nível Sobre a Inteligência Artificial. (2019). *Orientações éticas para uma IA de confiança*. Comissão Europeia.
- IBM. (s.d.). *Think – What is machine learning (ML)?* Consultado de
<https://www.ibm.com/topics/machine-learning>
- Lameira, J. S. (2021, 21-22 de outubro). *Sessão de encerramento proferida pelo Vice-Presidente do CSM, Juiz Conselheiro José Sousa Lameira* [Comunicação]. XV Encontro Anual do CSM (Conselho Superior de Magistratura). A (des)humanização da Justiça – Tecnologia como meio e não como fim, Beja, Portugal.
<https://www.csm.org.pt/xv-encontro-anual-do-csm/>
- Lusa (2023a, 17 de fevereiro). Governo avança com estratégia de inteligência artificial para a Justiça. *ADVOCATUS*.
<https://eco.sapo.pt/2023/02/17/governo-avanca-com-estrategia-de-inteligencia-artificial-para-a-justica/>
- Lusa (2023b, 23 de dezembro). Programa com IA promete “nova era” nas decisões da gestão penitenciária. *Observador online*.
<https://observador.pt/2023/12/23/programa-com-ia-promete-nova-era-nas-decisoes-da-gestao-penitenciaria/>
- Maranhão, J., Florêncio, J., & Almada, M. (2021). Inteligência artificial aplicada ao direito e o direito da inteligência artificial. *SUPREMA – Revista de Estudos Constitucionais*, 1(1), 154-180.
- Martins, J. (2021). Dilemas éticos e jurídicos do uso da inteligência artificial na prática jurídica. *RJLB*, (4), 919-952.
- Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: what it is and how it works. *AI and Society*, 39, 1871-1882.
<https://doi.org/10.1007/s00146-023-01635-y>
- Nunes, D., & Marques, A. (2018). Inteligência artificial e direito processual: vieses algoritmos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, 43(285), 421-447.
- OECD, Organisation for Economic Cooperation and Development. (2023). Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI. *OECD Digital Economy Papers*, (349), 4-72.
- PE, Parlamento Europeu. (2021a). *Relatório sobre a inteligência artificial no direito penal e a sua utilização pelas autoridades policiais e judiciárias em casos penais (2020/2016(INI)) – Comissão das Liberdades Cívicas, da Justiça e dos Assuntos Internos*.

- PE, Parlamento Europeu. (2021b). *Resolução sobre A inteligência artificial no direito criminal e a sua utilização pelas autoridades policiais e judiciárias em questões criminais*.
- Pedroso, J. (2002). Percurso(s) da(s) reforma(s) da administração da justiça – uma nova relação entre o judicial e o não judicial. *Oficina do CES Working paper*, (171), 1-43.
- Pedroso, J. (2006). A justiça em Portugal entre a(s) crise(s) e a(s) oportunidade(s) – Contributo para a construção de um novo paradigma de política pública de justiça. *Scientia Iuridica*, 55(306), 263-302.
- Pedroso, J. (2021). Pandemia e Tribunais – neutralização, ausência e metamorfose digital. In Á. Garrido, & H. Augusto Costa (Orgs.), *Um vírus que nos (re)une: reflexões da FEUC* (pp. 179-184). Vida Económica.
- Pedroso, J., & Dias, J. P. (2004). As crises e as reformas da administração da justiça. *JANUS 2004 – Anuário de Relações Exteriores*, (8), 184-185.
<http://hdl.handle.net/10316/33084>
- PORTUGAL.GOV.PT. (2023, 17 de fevereiro). Lançada Estratégia Govtech centrada em inteligência artificial e tecnologias emergentes para transformar a Justiça.
<https://www.portugal.gov.pt/pt/gc23/comunicacao/noticia?i=lancada-estrategia-a-govtech-centrada-em-inteligencia-artificial-e-tecnologias-emergentes-para-transformar-a-justica>
- ProPublica. (s.d). *Sample-COMPAS-Risk-Assessment-COMPAS-"CORE"*.
<https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html>
- Reiling, D. (2009). *Technology for Justice How Information Technology Can Support Judicial Reform*. Leiden University Press.
- Reiling, D. (2020). Courts and Artificial Intelligence. *International Journal for Court Administration*, 11(2), 1-10.
- Revista JUSTICE TRENDS. (s.d.). *HORUS 360 iOMS – Sistema Inteligente de Gerenciamento de Infratores*.
<https://justice-trends.press/pt/projeto-horus-360-ioms/>
- Rodrigues, A. M. (2023). Justiça penal e inteligência artificial – uma justiça fitness? In C. Aranguena Fanego, M. de Hoyos Sancho, & E. G. Pillado (Dirs.), *El proceso penal ante una nueva realidad tecnológica europea* (pp. 207-230). Thomson Reuters Aranzadi.
- Santos, B. S. (2005). Os tribunais e as novas tecnologias de comunicação e de informação. *Sociologias*, (13), 82-109.
- SGMJ, Secretaria-Geral do Ministério da Justiça. (s.d.). *Guias – Guia Prático da Justiça*. JUSTIÇA.GOV.PT. <https://justica.gov.pt/Guias/Guia-Pratico-da-Justica>
- Silva, G. B. P., & Ehrhardt Júnior, M. (2020). Diretrizes éticas para a Inteligência Artificial confiável na União Europeia e a regulação jurídica no Brasil. *Revista IBERC*, 3(3), 1-28.
- Sourdin, T. (2018). Judge v Robot? Artificial Intelligence and Judicial Decision-Making. *UNSW Law Journal*, 41(4), 1114-1132.
- Surden, H. (2019). Artificial Intelligence and Law: An Overview. *Georgia State University Law Review*, 35(4), 1305-1337.

- Thompson, D. (2015). Creating New Pathways to Justice Using Simple Artificial Intelligence and Online Dispute Resolution. *International Journal of Online Dispute Resolution*, 1(2), 4-53. <https://ssrn.com/abstract=2696499>
- van Eijk, G. (2017). Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality. *Punishment & Society*, 19(4), 463-481. <https://doi.org/10.1177/1462474516666282>
- Wallace, A. (2017). From the Editor: The impact of technology on courts. *International Journal for Court Administration*, 8(2), 1.
- Wisser, L. (2019). Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing. *Georgetown Law – American Criminal Law Review*, 56(4), 1811-1832.
- Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum – Journal of the Academy of European Law*, 20, 567-583.
- Zhang, J., & Han, Y. (2022). Algorithms Have Built Racial Bias in Legal System-Accept or Not? *Advances in Social Science, Education and Humanities Research*, 631, 1217-1221.

Data de submissão: 30/06/2023 | **Data de aceitação:** 18/06/2024

