

Estrasburgo, 19 de dezembro de 2025

CEPEJ(2025)18 Final

**COMISSÃO EUROPEIA PARA A EFICIÊNCIA DA JUSTIÇA
(CEPEJ)**

**DIRETRIZES SOBRE A UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA
NOS TRIBUNAIS**

*Adotadas na 45.^a reunião plenária da CEPEJ
(Estrasburgo, 4 e 5 de dezembro de 2025).*

Índice

SUMÁRIO EXECUTIVO	3
1. INTRODUÇÃO	5
1.1. Preâmbulo	5
1.2. IA Generativa.....	5
1.3. A IA generativa no contexto do setor da justiça	6
2. AS DIRETRIZES.....	9
2.1. Aspectos operacionais para IA gerativa nos tribunais	9
2.1.1 Começar por identificar o problema, não a solução técnica	9
2.1.2. Identificar áreas adequadas para o uso IA de generativa.....	10
2.1.2.1. Gestão de processos e administração judicial.....	10
2.1.2.2. Gestão de documentos e processamento de informação	10
2.1.2.3. Serviços de informação e apoio ao utilizador.....	10
2.1.2.4. Apoio à redação e documentação judicial.....	10
2.1.3. Esteja ciente da existência de aplicações inadequadas	11
2.1.4. Privilegie ferramentas personalizadas seguras, em vez de soluções genéricas e padronizadas	12
2.1.5. Garantir uma gestão sólida na implementação de sistemas de IA generativa.....	14
2.1.6. Ter consciência da forma como as partes utilizam a IA generativa.....	16
2.1.7. Assegurar a sensibilização e formação dos utilizadores de IA	17
2.2. Aspectos normativos da IA generativa nos tribunais	18
2.2.1. Direito a um julgamento justo e independência judicial (artigo 6.º da CEDH)	19
2.2.1.1. Garantias de um julgamento justo.....	19
2.2.1.2. Princípio da legalidade, segurança jurídica e previsibilidade	19
2.2.1.3. Exclusividade da função jurisdicional.....	20
2.2.1.4. Responsabilidade humana obrigatória e supervisão	20
2.2.1.5. Independência judicial e natureza não vinculativa da IA.....	20
2.2.1.6. Transparência e fundamentação adequada.....	21
2.2.1.7. Direito de acesso a um tribunal.....	21
2.2.2. Direito à liberdade e à segurança (artigo 5.º da CEDH).....	21
2.2.3. Direito à vida privada e à proteção de dados (artigo 8.º da CEDH)	22
2.2.4. Direito a um recurso efetivo (artigo 13.º da CEDH)	22
2.2.5. Proibição da discriminação (artigo 14.º da CEDH e Protocolo n.º 12)	23
2.2.5.1. Princípio da não discriminação	23
2.2.5.2. Neutralidade e prevenção ativa de enviesamentos	23
2.2.6. Salvaguardas adicionais	23
2.2.6.1. Explicabilidade e fundamentação transparente.....	23
2.2.6.2. Proporcionalidade e cautela na implementação da IA generativa.....	24
2.2.6.3. Responsabilidade do Estado por danos causados pela utilização de IA generativa.....	24
3. GLOSSÁRIO	25

SUMÁRIO EXECUTIVO

As presentes Diretrizes estabelecem um quadro de referência para a implementação da inteligência artificial (IA) generativa na administração da justiça, com especial incidência na salvaguarda da segurança jurídica, da independência judicial, assegurada por uma supervisão humana efetiva, bem como da transparência e da rastreabilidade dos processos.

As ferramentas de uso geral atualmente disponibilizadas por grandes operadores tecnológicos constituem apenas uma etapa inicial, devendo a sua utilização evoluir para soluções especificamente concebidas para o setor da justiça. Sempre que adequado, deverá privilegiar-se o recurso a modelos linguísticos de grande dimensão implementados localmente ou a soluções que garantam um controlo soberano sobre os dados e as infraestruturas. Em qualquer caso, as ferramentas adotadas devem ser ajustadas às necessidades dos processos judiciais e utilizadas sob controlo público.

A implementação da inteligência artificial generativa na administração da justiça deve respeitar plenamente a Convenção Europeia dos Direitos do Homem (CEDH), a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial (CETS 225), a Convenção 108+ sobre a Proteção das Pessoas relativamente ao Tratamento Automatizado de Dados de Caráter Pessoal, bem como a Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente.

Estas Diretrizes assentam em princípios fundamentais, designadamente: o exercício da função jurisdicional pertence exclusivamente aos tribunais; o acesso efetivo a um juiz humano deve ser sempre garantido; os resultados produzidos por sistemas de inteligência artificial generativa não têm carácter vinculativo; e a utilização dessas tecnologias no contexto judicial, incluindo na preparação das decisões, deve ser plenamente transparente. As exigências em matéria de proteção da vida privada impõem a pseudonimização dos dados utilizados e o controlo exclusivo dos dados e das infraestruturas pela autoridade pública competente responsável pela prestação de serviços de tecnologias da informação ao sistema judicial. É fundamental considerar a legalidade, a independência judicial, a não discriminação, o direito a um recurso eficaz e a rastreabilidade e explicabilidade do raciocínio.

É preconizada uma abordagem de implementação progressiva, assente em projetos-piloto, métricas de avaliação e na possibilidade de reversão, sempre que necessário. Esta abordagem é sustentada pelo princípio da subsidiariedade, segundo o qual devem ser identificadas, em primeiro lugar, as necessidades funcionais e consideradas prioritariamente soluções tecnológicas convencionais,

recorrendo-se à inteligência artificial generativa apenas quando essas soluções se revelem inadequadas ou ineficientes.

A utilização da inteligência artificial generativa pode revelar-se particularmente útil em diversos contextos profissionais, nomeadamente na gestão processual, no tratamento documental, nos serviços de informação ao público e no apoio à redação de atos judiciais de natureza rotineira. Todavia, sempre que essas tecnologias sejam utilizadas para apoiar as autoridades judiciais na análise dos factos, na interpretação e aplicação do direito ou na apreciação de situações concretas, devem ser observados requisitos especialmente rigorosos. Para garantir uma implementação bem-sucedida, é essencial realizar avaliações *ex ante* exaustivas dos riscos e impactos antes de qualquer implementação. Além disso, a implementação de mecanismos contínuos de monitorização e *feedback* permitirá que tanto o tribunal como a administração da justiça façam ajustes e melhorias. Por fim, a formação de todo o pessoal do tribunal e dos utilizadores do sistema é crucial para garantir a utilização eficaz do sistema.

Por último, o Estado deve continuar a ser responsável por quaisquer danos decorrentes da utilização de sistemas de IA generativa na administração da justiça, garantindo que os utilizadores mantenham vias eficazes de recurso e reparação em caso de prejuízo.

1. INTRODUÇÃO

1.1. Preâmbulo

1. Estas diretrizes têm como objetivo dar conselhos práticos sobre como implementar com segurança a Inteligência Artificial (IA) generativa na administração da justiça. Elas são dirigidas às autoridades responsáveis pela administração da justiça e aos profissionais do setor judicial. Devem ser revistas e atualizadas periodicamente, de acordo com os recentes desenvolvimentos tecnológicos.

2. O uso de IA na justiça deve respeitar os direitos garantidos pela Convenção Europeia dos Direitos Humanos (CEDH), a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial e Direitos Humanos, Democracia e Estado de Direito (CETS225), a proteção de dados pessoais à luz da Convenção do Conselho da Europa para a Proteção de Dados Pessoais (CETS108), bem como os princípios estabelecidos na Carta Ética Europeia sobre o Uso de Inteligência Artificial em Sistemas Judiciais e os seus Ambientes (Carta da CEPEJ) e, quando aplicável, o Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024 que estabelece regras harmonizadas sobre a inteligência artificial (Regulamento da IA).

3. Estas diretrizes foram elaboradas pelo Grupo de Trabalho da CEPEJ sobre Ciberjustiça e IA (CEPEJ-GT-CYBERJUST) com base na contribuição de especialistas.¹ Elas levam em consideração trabalhos preliminares da CEPEJ, tais como a «Nota informativa sobre o uso da IA generativa por profissionais do setor judicial em contexto profissional», as «Reflexões do Conselho Consultivo da CEPEJ sobre Inteligência Artificial sobre como fazer com que a Inteligência Artificial funcione para o poder judicial» e o «1.º Relatório sobre a utilização da Inteligência Artificial (IA) no poder judicial, com base nas informações contidas no Centro de Recursos da CEPEJ sobre Ciberjustiça e IA» e são ilustradas através de ferramentas apresentadas no Centro de Recursos da CEPEJ sobre Ciberjustiça e IA.

1.2. IA Generativa

4. A IA generativa é um ramo da IA que se concentra em ensinar máquinas a criar novos conteúdos que imitam de forma convincente as produções humanas. Em vez de apenas classificar ou prever dados, estes sistemas aprendem com grandes

¹ Javier Ercilla Garcia, Juiz, Espanha

conjuntos de dados para identificar padrões e estruturas subjacentes, permitindo-lhes gerar texto, imagens, música, vídeos e outros tipos de conteúdo.²

5. No campo da geração de texto, a evolução das ferramentas de «processamento de linguagem natural» permitiu que softwares automatizados interpretassem e gerassem linguagem naturalmente. Anteriormente, a compreensão da linguagem humana representava um desafio para os softwares de IA devido à sua ambiguidade e complexidade, mas com os avanços no *deep learning*, os modelos tornaram-se capazes de aprender os padrões da nossa linguagem e produzir respostas ou textos coerentes. Isso significou uma mudança de paradigma na forma como as máquinas processam e geram conteúdo a partir de conteúdo textual.

6. A tecnologia por detrás da IA generativa baseia-se em aprendizagem profunda e arquiteturas avançadas de redes neuronais. Essencialmente, estes modelos conhecidos como *Large Language Models* ou *LLMs*, são treinados com grandes volumes de dados para identificar padrões complexos (quer sequências de palavras, pixéis de imagem ou notas musicais) o que permite gerar novo conteúdo que imita as características do material original.

7. Apesar da sua eficácia numa ampla gama de tarefas e adoção em *chatbots*, a IA generativa apresenta atualmente limitações, entre as quais se encontram as “alucinações”. Em geração de linguagem natural, uma alucinação é um resultado que o modelo produz sem suporte fiel na entrada ou conhecimento de referência: pode mesmo contradizer explicitamente o texto fonte (intrínseco) ou adicionar dados falsos ou impossíveis (extrínseco). Essencialmente, é informação inventada que o sistema apresenta como se fosse um facto.³

8. Outras limitações dignas de nota dizem respeito a potenciais questões de direitos autorais decorrentes dos dados de treino originais desses modelos, bem como à divulgação não solicitada de dados de entrada a terceiros e sua reutilização para treinar o modelo de IA generativa, potencialmente violando a regulamentação aplicável em matéria de proteção de dados.

1.3. A IA generativa no contexto do setor da justiça

9. A utilização de soluções de IA generativa por profissionais da justiça está a aumentar, seja através da geração de resumos, correção de texto, redação de fundamentos jurídicos com base na jurisprudência, interação com os litigantes

² Warankar, M., & Patil, R. (2024). Generative Artificial Intelligence. International Journal of Scientific Research in Engineering and Management, 8(04). doi:10.55041/IJSREM31146

³ Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2022). Survey of hallucination in natural language generation. arXiv. <https://arxiv.org/abs/2202.03629>

(explicar as sentenças aos clientes), etc. De acordo com o relatório *Future Ready Lawyer 2024*⁴, em que foram inquiridos 712 profissionais do direito nos Estados Unidos e nove países europeus, 68% dos escritórios de advogados e 76% dos departamentos jurídicos das empresas utilizam IA generativa pelo menos uma vez por semana, com mais de um terço a utilizar diariamente (35% nos departamentos jurídicos das empresas e 33% nos escritórios de advogados). Na Europa, os Países Baixos lideram a utilização diária de IA generativa com 42%, seguidos pela Alemanha com 38%. A principal utilização destas ferramentas é para a automatização de documentos (67%), gestão do tempo (66%) e planeamento estratégico (62%). Além disso, 58% dos escritórios de advocacia planeiam aumentar o investimento em IA nos próximos três anos, assim como 73% dos departamentos jurídicos das empresas.

10. De acordo com um inquérito realizado entre os membros da Rede Europeia de Ciberjustiça da CEPEJ em outubro de 2025, 46% dos inquiridos⁵ confirmaram a utilização de IA generativa nos tribunais, dos quais 74% se referem a aplicações padrão/prontas a usar (por exemplo, Microsoft Copilot, Chat GPT, DeepL) e 52% a aplicações personalizadas/adaptadas (ambas as opções possíveis). As utilizações mais mencionadas da IA generativa dizem respeito à síntese de documentos e outras funções de apoio editorial. Nenhum inquirido mencionou a utilização da IA generativa para a estimativa quantitativa de resultados jurídicos (ver Figura 1).

⁴ Wolters Kluwer. (2024). 2024 Future Ready Lawyer Survey Report: Legal innovation: Seizing the future or falling behind? Wolters Kluwer Legal & Regulatory. <https://www.wolterskluwer.com/en/know/future-ready-lawyer-2024>

⁵ n=33 inquiridos de 43 Estados-Membros do Conselho da Europa. Decimais arredondados para números inteiros.

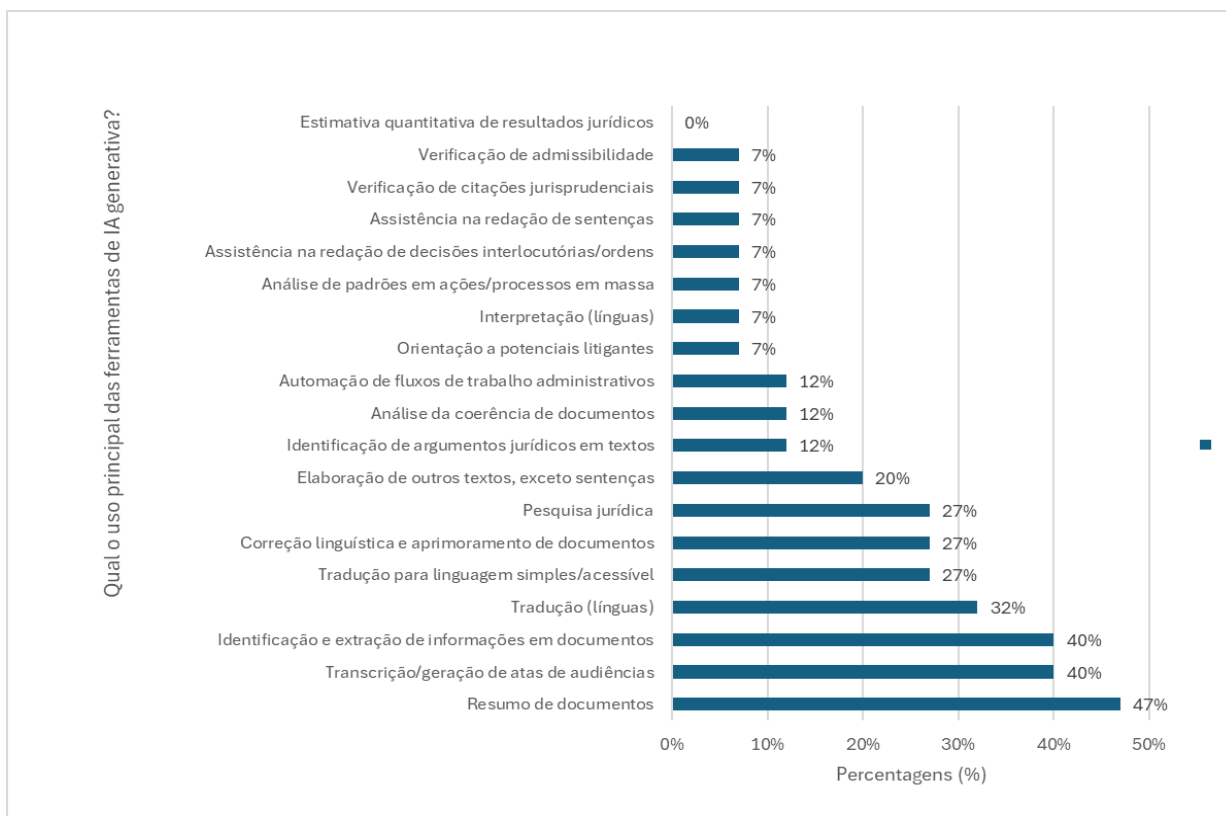


Figura 1: Utilização de IA gerativa nos tribunais.

11. Vários países estão a facilitar o uso de ferramentas de IA generativa na administração pública e nos tribunais por meio de acordos de confidencialidade, proteção de dados de uso e garantias de que tais dados não serão reutilizados para *retraining* de modelos^{6 7}.

12. No campo jurídico, o uso da IA generativa deu origem a processos judiciais com citações jurídicas falsas^{8 9}, o que tem por vezes levado a sanções para os seus utilizadores¹⁰. Uma vez que as “alucinações” são inevitáveis¹¹ ou, no melhor

⁶ Artificial Lawyer. (2025, April 24). UK Courts roll out Microsoft Copilot for judges, update GenAI rules. <https://www.artificiallawyer.com/2025/04/24/uk-courts-roll-out-microsoft-copilot-for-judges-update-genai-rules/>

⁷ El País (2025, 22 September). Do you already know what carpeta.jusitcia.es is? Retrieved from <https://cincodias.elpais.com/extras/2025-09-22/ya-sabes-lo-que-es-carpetajusticiaes.html>

⁸ Weiser, B. (2023, May 27). Here’s what happens when your lawyer uses ChatGPT. The New York Times. <https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>

⁹ MIT Sloan Teaching & Learning Technologies. (2024, November 12). When AI gets it wrong: Addressing AI hallucinations and bias. <https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/>

¹⁰ Carrick, D., & Kesteven, S. (2023, June 24). Este advogado norte-americano utilizou o ChatGPT para pesquisar um resumo jurídico, com resultados embaraçosos. Todos nós podemos aprender com o seu erro. ABC News. <https://www.abc.net.au/news/2023-06-24/us-lawyer-uses-chatgpt-to-research-case-with-embarrassing-result/102490068>

¹¹ Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is Inevitable: An Innate Limitation of Large Language Models. arXiv. <https://arxiv.org/abs/2401.11817>

das hipóteses, passíveis de mitigação ¹², de acordo com o atual estado da arte, é um risco a considerar ao usar IA gerativa na justiça, sem as salvaguardas adequadas de supervisão humana.

2. AS DIRETRIZES

2.1. Aspectos operacionais para IA gerativa nos tribunais

2.1.1 Começar por identificar o problema, não a solução técnica

13. O princípio da neutralidade tecnológica estabelece que as administrações públicas devem concentrar-se nas necessidades funcionais e, com base nisto, selecionar soluções tecnológicas que podem ser adaptadas ao longo do tempo. Esta abordagem visa minimizar dependências tecnológicas, evitar impor implementações técnicas específicas ou produtos, e permanecer adaptável num ambiente tecnológico em rápida evolução. Este princípio deve estender-se à utilização de IA generativa na administração judicial através do **princípio da subsidiariedade**, em que as administrações judiciais primeiro identificam claramente as necessidades funcionais e, em segundo lugar consideram diferentes soluções tecnológicas, convencionais ou novas, antes de optarem pela mais adequada e menos arriscada. Isto não deve impedir os tribunais de procurarem melhorar continuamente a eficiência da justiça com o apoio da tecnologia quando apropriado, nem de se absterem de introduzir IA generativa em casos de utilização bem estabelecidos que cumprem necessidades funcionais bem identificadas.

14. Tanto os tribunais como as administrações da justiça enfrentam vários desafios, incluindo a escassez de pessoal, os elevados volumes de ações em massa, tarefas de redação repetitivas, a gestão ineficiente de ficheiros, dificuldades em encontrar jurisprudência, ou problemas na comunicação com os utilizadores do sistema judicial e falta de interoperabilidade entre os sistemas informáticos atualmente em uso.

15. Uma vez identificadas as necessidades e os problemas operacionais dentro da administração, o próximo passo é selecionar soluções técnicas adequadas, de acordo com os princípios da neutralidade tecnológica e da subsidiariedade da IA.

16. Ao selecionar as soluções técnicas mais adequadas, a relação custo eficiência deve ser considerada. Tal deve incluir fatores como o consumo de energia, medidas de cibersegurança, licenciamento, manutenção, desenvolvimento e custos de consultoria, para além do preço de compra inicial.

¹² Nie, F., Yao, J.-G., Wang, J., Pan, R., & Lin, C.-Y. (2019, July). A simple recipe towards reducing hallucination in neural surface realisation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 2673-2679). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1256>

2.1.2. Identificar áreas adequadas para o uso IA de generativa¹³

2.1.2.1. Gestão de processos e administração judicial

17. A IA generativa poderia realizar a gestão preditiva da carga de trabalho através da análise de padrões históricos e criar repositórios de jurisprudência organizados por tipo de caso e assunto. Estas ferramentas poderiam facilitar a identificação de padrões argumentativos em alegações semelhantes, permitindo a detecção de processos que podem ser resolvidos uniformemente quando suscitam a mesma questão legal, ou ajudando a alocar casos. A IA generativa poderia também ativar o reconhecimento automatizado de questões em disputa através da comparação de queixas e respostas.

2.1.2.2. Gestão de documentos e processamento de informação

18. A IA generativa poderia gerar automaticamente atas de audiências, elaborar índices temáticos e cronológicos para processos volumosos e extrair dados relevantes de documentos não estruturados. Da mesma forma, poderia ser usada para obter informações importantes, como nomes, datas e valores, incorporando-as automaticamente em modelos predefinidos.

2.1.2.3. Serviços de informação e apoio ao utilizador

19. A IA generativa podia permitir a criação de *chatbots* especializados por tema jurídico e sistemas de-orientação procedimental para cidadãos. Estas aplicações devem ser concebidas com especial atenção à exatidão da informação prestada e à clareza da comunicação com utilizadores não especializados, utilizando linguagem clara.

2.1.2.4. Apoio à redação e documentação judicial

20. A IA generativa poderia ser utilizada para a criação automática de documentos processuais padrão, para auxiliar na elaboração de ordens e na verificação da consistência dos documentos. Poderia gerar resumos estruturados de sentenças, documentos processuais e recursos, exigindo verificação judicial para garantir que o resumo corresponda fielmente ao original. Poderia também facilitar a preparação de documentos, como relatórios, esquemas, ou mapas mentais, que apresentam uma visão geral dos pontos de concordância e discordância.

21. Além disso, poderia ser usada para a análise de reclamações e respostas com sugestões de jurisprudência relacionada; identificação e síntese da jurisprudência relevante para cada caso; análise preliminar de recursos sobre a

¹³ A lista a seguir não é exaustiva.

conformidade com requisitos formais e técnicos; e fazer a ligação entre diferentes documentos processuais, produzindo documentos organizados nos quais a resposta do réu está sistematicamente associada com cada um dos argumentos do autor.

2.1.3. Esteja ciente da existência de aplicações inadequadas

22. O Regulamento da UE sobre IA (Regulamento (UE) 2024/1689) estabelece proibições específicas sobre práticas de IA que são particularmente relevantes para a administração da justiça e que possam ser IA generativa. O artigo 5.º do Regulamento estabelece as seguintes práticas de IA proibidas, que podem servir de ilustração também para países fora do seu âmbito de aplicação: Sistemas de IA que utilizam técnicas subliminares ou manipuladoras (artigo 5.º, n.º 1, alínea a)); Sistemas de IA que exploram vulnerabilidades (artigo 5.º, n.º 1, alínea b)); Sistemas de classificação social (artigo 5.º, n.º 1, alínea c)); Sistemas de avaliação de risco para a prevenção de crimes (Artigo 5.º, n.º 1, alínea d)); Sistemas de expansão de bases de dados de reconhecimento facial (Artigo 5.º, n.º 1, alínea e)); Sistemas de inferência de emoções (Artigo 5.º, n.º 1, alínea f)).

23. Além disso, embora não seja explicitamente proibido, o Regulamento da IA da UE classifica certos sistemas de IA no domínio da justiça como de alto risco nos termos do Anexo III, particularmente aqueles destinados a auxiliar as autoridades judiciais na pesquisa e interpretação de factos e leis, e na aplicação da lei a conjuntos concretos de factos. Esses sistemas exigem o cumprimento de requisitos rigorosos, mas não são totalmente proibidos.

24. Embora os LLMs possam ser úteis na recolha de provas na fase de investigação, eles não deveriam substituir a avaliação judicial de provas, na fase de julgamento. Embora existam vertentes da IA lógica e neuro-simbólica que integram modelos causais¹⁴, a IA generativa carece da capacidade de «teorização prospetiva e desenvolvimento de lógica causal»^{15 16}.

25. A tomada de decisões deve permanecer sob controlo humano. Casos que exijam uma interpretação normativa inovadora ou a análise de princípios jurídicos abstratos não devem ser confiados à IA, uma vez que esta carece de «compreensão intrínseca do que é na realidade verdadeiro ou falso»¹⁷. Da mesma forma, o uso de IA generativa deve ser evitado em questões que exijam a interpretação evolutiva de regulamentos recentes, uma vez que a IA generativa

¹⁴ Colelough, B. C., & Regli, W. (2025). Neuro-symbolic AI in 2024: A systematic review. arXiv. <https://arxiv.org/abs/2501.05435>

¹⁵ Felin, T., & Holweg, M. (2024). Theory Is All You Need: AI, Human Cognition, and Decision Making. Available at: <https://ssrn.com/abstract=4737265>.

¹⁶ Li, X., Cai, Z., Wang, S., Yu, K., & Chen, F. (2025). A survey on enhancing causal reasoning ability of large language models. arXiv. <https://arxiv.org/abs/2503.09326>

¹⁷ Ibid.

pode apresentar limitações na adaptação da sua compreensão a regulamentos inovadores se não tiver sido especificamente treinada com eles¹⁸.

26. A análise da constitucionalidade ou conformidade com tratados internacionais requer um nível de interpretação axiológica que ultrapassa as capacidades atuais da IA. Mesmo que este não fosse o caso no futuro, a questão só pode ser deixada para o julgamento humano, visto que afeta a forma como os cidadãos decidiram governar-se a si próprios. Portanto, o "zeitgeist" de cada um momento irá influenciar o "controlo da convencionalidade e constitucional" e este não pode ser deixado ao critério de uma IA.

2.1.4. Privilegie ferramentas personalizadas seguras, em vez de soluções genéricas e padronizadas

27. A distinção entre ferramentas comerciais genéricas de IA generativa e ferramentas judiciais de IA generativa personalizadas é altamente relevante para a administração da justiça porque implica não apenas especificações técnicas, mas questões fundamentais de proteção de dados, independência judicial e integridade do processo jurídico.

28. As ferramentas comerciais gratuitas de IA generativa apresentam riscos significativos. Quando os profissionais judiciais utilizam serviços de IA generativa não oficiais, soluções comerciais gratuitas ou pagas, correm o risco de incorrer inadvertidamente numa transferência de dados internacional, principalmente para centros de dados nos EUA, alojados nas empresas que fornecem estes serviços. Estes dados subsequentemente tornam-se disponíveis para fins de *retraining* do modelo, removendo-os efetivamente do controlo administrativo judicial. Tais práticas violam os princípios fundamentais de proteção de dados estabelecidos na Convenção 108+¹⁹ e criam vulnerabilidades inaceitáveis no ecossistema de informação judicial.

29. As soluções comerciais genéricas de IA generativa, mesmo quando oferecidas por meio de assinaturas pagas, normalmente carecem do treinamento especializado necessário para o raciocínio jurídico e para a documentação judicial. Essas ferramentas, projetadas para uso geral do consumidor, demonstram compreensão limitada da terminologia jurídica, dos requisitos processuais e dos padrões de raciocínio subtis característicos do trabalho judicial. A ausência de integração com bases de dados jurídicas verificadas aumenta o risco de «alucinações», ou seja, a geração de disposições jurídicas

¹⁸ Ercilla García, J. (2025). Justicia automatizada: entre las inteligencias artificiales que fingen y las que persuaden. *Lex Social, Revista De Derechos Sociales*, 15 (1), 1-39. <https://doi.org/10.46661/lexsocial.11652>

¹⁹ Conselho da Europa. (2018). Convenção para a Proteção das Pessoas no que diz respeito ao Tratamento Automatizado de Dados de Caráter Pessoal [Convenção 108 Modernizada], Série Tratados do Conselho da Europa - No. 223. Estrasburgo

inexistentes ou de jurisprudência deturpada, o que representa uma grave ameaça à segurança jurídica.

30. Em caso de recorrer a soluções comerciais, a implementação de IA generativa na administração da justiça deve ser exclusivamente mediada através de acordos contratuais formais entre autoridades públicas e prestadores de serviços de IA. Em caso de recorrer a soluções internas, a implementação de IA generativa na administração da justiça deve ser mediada através de políticas sobre a utilização da IA na administração da justiça²⁰ ou um mais amplo sistema de segurança da informação²¹. Estes contratos ou políticas devem estabelecer estruturas abrangentes abordando:

- Protocolos rigorosos de governança de dados, garantindo que não haja transferência não autorizada para terceiros
- Procedimentos obrigatórios de destruição de dados após períodos de retenção especificados
- Proibições explícitas contra o uso de dados judiciais para *retraining* de modelos sem autorização expressa
- Delimitação clara da responsabilidade por violações de dados ou falhas do sistema
- Conformidade com os regulamentos nacionais e europeus de proteção de dados
- Auditorias de segurança regulares e requisitos de testes de penetração
- Procedimentos transparentes de resposta a incidentes
- Mecanismos para o envolvimento contínuo das partes interessadas (por exemplo, consultas públicas, comissões de supervisão, avaliações periódicas)

31. Modelos especializados de IA generativa judicial representam um requisito crítico para a implementação eficaz da IA em contextos jurídicos. A linguagem jurídica constitui um domínio técnico com terminologia especializada que muitas vezes diverge significativamente do uso comum. Termos como «consideração», «prejuízo», «ordem pública» ou «boa-fé» têm significados jurídicos específicos que diferem substancialmente das suas definições quotidianas. Essa complexidade semântica requer modelos de IA generativa treinados especificamente em *corpora* jurídicos de cada jurisdição, garantindo o uso preciso da terminologia jurídica técnica e a interpretação contextualmente correta.

32. As soluções personalizadas de IA generativa judicial podem incorporar:

- Integração com bases de dados jurídicas oficiais e fontes jurisprudenciais verificadas

²⁰ Por exemplo, Comissão Técnica Estatal para a Administração Judicial Eletrónica (2024) da Espanha. Política sobre o uso da inteligência artificial na administração da justiça. Pode ser encontrada uma tradução para o inglês em: <https://www.administraciondejusticia.gob.es/cteaje/normativa-complementaria>.

²¹ Por exemplo, o Decreto Real Espanhol 311/2022, que regulamenta o Esquema de Segurança Nacional.

- Formação específica sobre jurisdição em textos jurídicos, decisões judiciais e documentos processuais
- Capacidades multilingues que refletem a diversidade linguística dos sistemas judiciais europeus
- Módulos especializados para diferentes áreas do direito (civil, penal, administrativo, constitucional)
- Sistemas integrados de verificação de citações com referências cruzadas a registos jurídicos oficiais
- Detecção e sinalização automatizadas de potenciais inconsistências jurídicas ou anacronismos

33. A soberania da infraestrutura constitui uma consideração essencial. Os sistemas judiciais de IA generativa devem operar em infraestruturas sob o controlo exclusivo das autoridades judiciais ou de entidades governamentais quando se opta por modelos de código aberto executados em servidores públicos, ou em colaboração com empresas fornecedoras quando se opta por modelos comerciais.

2.1.5. Garantir uma gestão sólida na implementação de sistemas de IA generativa

34. A implementação de sistemas de IA generativa na administração judicial deve seguir uma abordagem abrangente que inclua vários elementos-chave: uma estratégia de implementação incremental com processos adequados de gestão da mudança e fases-piloto com avaliações para revisão, consideração cuidadosa dos modelos de licenciamento (tecnologia proprietária versus tecnologia de código aberto), suporte contínuo do fornecedor e requisitos de formação, auditorias regulares, monitorização e revisão do desempenho do sistema de IA e riscos associados.

35. Para garantir uma melhor compreensão da tecnologia, bem como para reduzir potenciais barreiras à adoção e facilitar a gestão da mudança, sugere-se que a implementação da IA generativa na administração da justiça siga uma abordagem gradual e passo a passo, diferenciada de acordo com a complexidade e a natureza das questões abordadas:

Fase Preparatória: Antes de qualquer implementação, devem ser realizadas auditorias de preparação institucional, deve ser ministrada formação básica ao pessoal, devem ser estabelecidos protocolos de segurança e devem ser criados comités de supervisão.

Primeira Fase: Limitado a funções administrativas e de apoio básicas que não afetam diretamente a tomada de decisões judiciais. Esta fase deve centrar-se na gestão de documentos e no processamento de informações, tais como a extração de dados relevantes de documentos não

estruturados, a obtenção de informações essenciais (nomes, datas, montantes) e a sua incorporação automática em modelos predefinidos. Da mesma forma, poderia incluir a gestão básica da carga de trabalho através da análise de padrões históricos e da assistência na atribuição de processos. Também podem ser implementados serviços básicos de informação e assistência ao utilizador, tais como sistemas de orientação processual em linguagem simples para os utilizadores e *chatbots* especializados para consultas processuais simples. Estes devem ser complementados com uma declaração de exoneração de responsabilidade sobre os riscos envolvidos. A transcrição não simultânea das audiências também pode ser considerada nesta fase. Tanto os tribunais como a administração da justiça devem avaliar o impacto na carga de trabalho, monitorizar a satisfação dos utilizadores e medir o tempo de adaptação.

Segunda Fase: Expansão para funções mais sofisticadas de apoio analítico e de redação. Esta fase deve abranger capacidades avançadas de gestão de documentos e processamento de informações, incluindo a geração automática de atas de audiências e a ligação de argumentos entre diferentes escritos processuais. Esta fase também pode estender-se ao apoio à redação e à documentação judicial, com a criação automática de documentos processuais padrão, assistência na redação de ordens de rotina e verificação da consistência dos documentos. Da mesma forma, com a criação de repositórios de jurisprudência organizados por tipo de caso e assunto e o reconhecimento automatizado de questões controversas através da comparação de reclamações e respostas, bem como a elaboração de índices temáticos e cronológicos para processos volumosos. A análise da jurisprudência e o apoio à decisão devem ser introduzidos para a análise preliminar dos recursos relativos ao cumprimento dos requisitos formais e técnicos. Os tribunais devem realizar testes A/B para comparação de eficiência, avaliar o viés das decisões e analisar os padrões de utilização.

Terceira Fase: Introdução de funções analíticas e consultivas mais avançadas, mantendo uma supervisão judicial rigorosa. Esta fase requer uma avaliação positiva das fases anteriores, aprovação específica para cada subfase e protocolos de reversão claramente definidos. Esta fase pode ser subdividida em três subfases:

Subfase inicial: Focada na pesquisa de jurisprudência e na organização de documentos através da análise de reclamações e respostas com sugestões de jurisprudência relacionada, e identificação e organização sistemática da jurisprudência relevante para cada caso para consideração judicial. O apoio à redação deve incluir a geração de resumos estruturados de sentenças,

documentos processuais e recursos, exigindo uma verificação judicial abrangente. Funções avançadas de gestão de processos podem facilitar a identificação de padrões argumentativos em reclamações semelhantes e a detecção de processos que possam levantar questões jurídicas semelhantes.

Subfase intermédia: Pesquisa jurisprudencial aprimorada e organização de documentos para apoiar a preparação judicial, fornecendo precedentes jurídicos sistematicamente organizados e jurisprudência relevante para consideração judicial. O apoio à redação pode incluir a preparação de documentos preparatórios, tais como relatórios, esboços e mapas mentais que apresentem uma visão geral dos pontos de concordância e discordância entre as partes.

Subfase avançada: Em procedimentos altamente padronizados com precedentes jurídicos bem estabelecidos, a IA generativa poderia ajudar fornecendo análises abrangentes da jurisprudência com reconhecimento de padrões em casos semelhantes, identificando precedentes processuais e potenciais inconsistências para revisão judicial e propondo redações que incorporem a jurisprudência relevante. Da mesma forma, poderia ajudar a elaborar modelos estruturados para documentos processuais de rotina com base em práticas judiciais estabelecidas, enquanto a análise e a tomada de decisões humanas obrigatórias permaneceriam em vigor para todas as determinações legais e judiciais e a aprovação final de cada resultado.

36. Todas as medidas acima referidas devem ser aplicadas como projetos-piloto em procedimentos de ordens de pagamento, reclamações avaliadas por montantes determinados, julgamentos verbais de pequenos montantes, multas administrativas normalizadas ou reclamações laborais por montantes incontroversos, em que todas as provas são documentais.

37. As questões que afetam os direitos fundamentais, envolvem menores ou pessoas incapacitadas, ou implicam medidas de privação de liberdade exigem a intervenção judicial humana plena, limitando a IA generativa a funções de apoio técnico.

2.1.6. Ter consciência da forma como as partes utilizam a IA generativa

38. Os tribunais e a administração da justiça devem estabelecer protocolos claros para a monitorização e gestão da utilização de IA generativa por advogados e partes nos processos. Tal pode incluir a exigência de que as partes e os seus

representantes legais declarem quando tenham recorrido a sistemas de IA generativa na preparação de documentos jurídicos, na investigação do processo ou na recolha de prova, quando tal seja considerado adequado.

39. As regras processuais devem prever a obrigação de notificar a utilização de IA generativa na preparação do processo, com sanções adequadas em caso de omissão dessa declaração ou de dependência excessiva de conteúdos gerados por IA sem a devida verificação.

40. Os tribunais devem estar particularmente atentos a múltiplas formas de conduta imprópria relacionada com a IA, incluindo a apresentação de prova fabricada, citações jurídicas inexistentes ou conteúdos gerados por IA que não tenham sido devidamente revistos por profissionais qualificados; tentativas de manipulação de sistemas de IA através de *prompts* ocultos ou instruções incorporadas nos documentos, como texto invisível, formatação invulgar ou instruções microscópicas destinadas a comprometer a análise assistida por IA; e a utilização de IA generativa para criar vídeos ou provas gráficas fraudulentas. Todas estas práticas devem ser reguladas tanto a nível processual como penal.

41. Para manter a equidade processual e a igualdade de direitos, os tribunais devem assegurar que todas as partes têm acesso igual à informação sobre a utilização de IA generativa nos processos e que nenhuma parte obtenha uma vantagem injusta através de assistência por IA não declarada.

2.1.7. Assegurar a sensibilização e formação dos utilizadores de IA

42. Juízes, magistrados do Ministério Público, funcionários administrativos e todo o pessoal judicial devem receber formação adequada para compreender, gerir e avaliar criticamente as ferramentas de IA generativa disponíveis, tanto as utilizadas internamente como as existentes no mercado. A formação especializada e uma maior consciencialização sobre o potencial e as limitações destas ferramentas assegurarão uma utilização ética e eficaz da tecnologia, preservando sempre o julgamento humano na tomada de decisões.

43. A implementação da IA na administração da justiça deve caracterizar-se pela transparência e pela automatização. Para garantir segurança jurídica e consistência, os juízes não devem ter de realizar operações específicas para ativar o sistema; a sua utilização deve estar previamente definida pela lei processual.

44. As competências nucleares dos juízes devem centrar-se na manutenção e aplicação do seu juízo crítico, uma qualidade inerente e exigida pela função judicial, avaliando rigorosamente todos os resultados fornecidos pela IA. Tal implica evitar uma postura complacente ou passiva face às sugestões tecnológicas e adotar, em vez disso, uma atitude crítica, e até dissidente,

relativamente a determinadas alternativas ou propostas geradas por estes sistemas.

45. Caso os juízes sejam dotados de sistemas de IA generativa com os quais possam interagir diretamente, devem receber formação especializada na sua utilização (engenharia de *prompts*, engenharia de contexto, entre outros). Devem adquirir uma compreensão técnica sólida das limitações inerentes a estes modelos: o alcance das suas capacidades, as áreas em que tendem a falhar, os seus pontos fortes e os tipos de questões que não são tecnicamente viáveis ou adequadas.

46. Os juízes devem ser capazes de analisar e questionar os resultados produzidos pela IA generativa, identificando potenciais enviesamentos, incoerências ou limitações no raciocínio. Devem desenvolver critérios para avaliar quando e como é apropriado utilizar a IA generativa em diferentes tipos de processos e adquirir a capacidade de transitar eficientemente entre fontes tradicionais (como bases de dados de jurisprudência e artigos doutrinários) e sistemas de IA generativa, de modo a obter a informação mais relevante e fiável.

47. Os juízes devem ser formados para evitar o chamado “efeito de bolha”, que surge quando, impressionado pela aparente eficiência e exatidão dos sistemas de IA generativa, o juiz desenvolve uma confiança excessiva na jurisprudência, nos relatórios ou nos resumos fornecidos, correndo o risco de abandonar a consulta independente das bases de dados originais de jurisprudência e de desconsiderar os documentos subjacentes.

48. Deve ser desenvolvida uma política de formação contínua para novas ferramentas de IA generativa à medida que estas surjam, bem como para os avanços que a própria IA venha a alcançar em diferentes domínios relevantes para a justiça (geração de imagens, vídeos, vozes, etc.). As instituições judiciais devem estabelecer mecanismos sistemáticos para manter os profissionais da justiça atualizados quanto às novas capacidades da IA e às suas potenciais implicações para a administração da justiça. Tal deve incluir a compreensão de novas formas de prova geradas por IA, a deteção de *deepfakes* e conteúdos sintéticos, e a adaptação a métodos evolutivos de manipulação digital que possam afetar os processos judiciais. Esta formação deve ser regularmente atualizada para refletir o estado da arte da tecnologia de IA generativa e as suas aplicações em contextos jurídicos. Para garantir consistência na aplicação, poderão ser considerados currículos de formação normalizados.

2.2. Aspetos normativos da IA generativa nos tribunais

49. A implementação de sistemas de IA generativa em contextos judiciais exige uma avaliação rigorosa à luz dos princípios jurídicos fundamentais consagrados nos quadros jurídicos europeus e nas disposições constitucionais nacionais. Para

assegurar o respeito por esses princípios jurídicos, podem ser utilizados o Instrumento de Avaliação da CEPEJ²² e a metodologia de análise de impacto HUDERIA²³.

2.2.1. Direito a um julgamento justo e independência judicial (artigo 6.º da CEDH)²⁴

2.2.1.1. Garantias de um julgamento justo

50. Os sistemas de IA generativa podem comprometer a equidade processual através de processos de tomada de decisão opacos, recomendações automatizadas não explicáveis ou vantagens processuais desiguais resultantes do acesso diferenciado das partes a ferramentas tecnológicas avançadas²⁵.

51. As salvaguardas técnicas decorrentes destas disposições devem incluir a implementação de sistemas de IA explicável²⁶, a rastreabilidade integral dos processos de raciocínio da IA²⁷, o acesso igual para todas as partes aos resultados gerados por IA e mecanismos técnicos que assegurem uma verdadeira igualdade de armas²⁸. Os sistemas de IA que apoiem decisões judiciais devem fornecer resultados demonstravelmente explicáveis e fiáveis, compreensíveis por especialistas não técnicos. Devem igualmente adotar medidas para minimizar imprecisões, assegurar a correção do raciocínio jurídico e proteger a segurança dos dados.

2.2.1.2. Princípio da legalidade, segurança jurídica e previsibilidade

52. Os sistemas de IA generativa que produzam “alucinações jurídicas” ou façam referência a normas jurídicas inexistentes representam riscos graves para a segurança jurídica e para o Estado de direito.

53. A implementação da IA não deve comprometer a previsibilidade e a coerência das decisões judiciais. As fontes de informação, os critérios e as metodologias utilizados pelos modelos de IA generativa devem ser integralmente

²² Comissão Europeia para a Eficácia da Justiça (CEPEJ). (4 de dezembro de 2023). Ferramenta de avaliação para a operacionalização da Carta Ética Europeia sobre a Utilização da Inteligência Artificial nos Sistemas Judiciais e no seu Ambiente. Conselho da Europa.

²³ Conselho da Europa, Comissão sobre Inteligência Artificial. (28 de novembro de 2024). Metodologia para a Avaliação de Riscos e Impactos dos Sistemas de Inteligência Artificial do Ponto de Vista dos Direitos Humanos, da Democracia e do Estado de Direito (Metodologia HUDERIA). Conselho da Europa.

²⁴ Convention for the Protection of Human Rights and Fundamental Freedoms, Nov. 4, 1950, ETS No. 5, 213 U.N.T.S. 221. https://www.echr.coe.int/documents/d/echr/convention_ENG

²⁵ O artigo 6.º, n.º 1, da CEDH garante o direito a «(...) um julgamento justo e público, num prazo razoável, por um tribunal independente e imparcial, estabelecido pela lei».

²⁶ Nível de compreensão de como o sistema baseado em IA chegou a um determinado resultado, consulte ISO/IEC TR 29119-11:2020

²⁷ Exigir que as organizações mantenham registos detalhados das operações de IA, incluindo entradas de dados, decisões de modelos, erros e interações dos utilizadores, consulte ISO/IEC 42001:2023

²⁸ A Ferramenta de Avaliação da CEPEJ identifica como risco n.º 8: «Risco de vantagem injusta para uma das partes no julgamento».

documentados e claramente justificados, a fim de evitar a introdução de incerteza ou arbitrariedade nos processos judiciais.

54. As salvaguardas técnicas devem incluir a validação automática de citações jurídicas, a utilização exclusiva de fontes jurídicas oficiais certificadas e sistemas de alerta para conteúdos gerados que não possam ser verificados com base em bases de dados jurídicas fidedignas²⁹.

2.2.1.3. Exclusividade da função jurisdicional

55. O poder jurisdicional pertence exclusivamente aos juízes, nos termos das disposições constitucionais e legais. A tomada de decisão, o raciocínio jurídico e a apreciação da prova não devem ser delegados em sistemas de IA generativa; tais sistemas só podem ser utilizados para tarefas auxiliares ou preparatórias que apoiem, mas nunca substituam, o juízo humano do órgão judicial.

2.2.1.4. Responsabilidade humana obrigatória e supervisão

56. Os juízes mantêm a responsabilidade última e exclusiva por todas as decisões judiciais, independentemente de qualquer assistência tecnológica utilizada no processo decisório. Os magistrados judiciais devem compreender plenamente as limitações de qualquer sistema de IA generativa utilizado e rever exaustivamente todos os resultados gerados por IA antes de os integrar em atos ou decisões judiciais.

2.2.1.5. Independência judicial e natureza não vinculativa da IA

57. A independência judicial³⁰ pode ser posta em causa quando os juízes sofrem pressão direta ou indireta para seguir recomendações geradas por IA, limitando a sua liberdade de julgamento, interpretação jurídica e autonomia decisória. Esta independência deve abranger a liberdade face à coerção tecnológica ou a enviesamentos sistemáticos introduzidos por recomendações da IA. Os juízes devem manter o direito de optar por não receber sugestões da IA.

58. As recomendações, sugestões ou resultados gerados por sistemas de IA nunca serão vinculativos para os juízes, nem a sua utilização será obrigatória, preservando-se plenamente a independência judicial na interpretação, fundamentação e decisão jurídica. Os juízes não devem ser obrigados a justificar especificamente o afastamento de propostas ou recomendações geradas por sistemas de IA.

²⁹ A Ferramenta de Avaliação da CEPEJ identifica como risco n.º 11: «Risco de geração e utilização de disposições legais inexistentes pela IA generativa».

³⁰ O artigo 6.º, n.º 1, da CEDH exige que o julgamento seja realizado por um «tribunal independente e imparcial estabelecido por lei».

59. A concepção dos modelos de IA generativa não deve privilegiar a maximização da sua capacidade persuasiva, de modo a minimizar o risco de influência psicológica indevida e de enviesamento por automatização que possa afetar negativamente a imparcialidade e a liberdade judicial na avaliação de pareceres judiciais gerados por IA³¹.

60. As salvaguardas técnicas devem ser proporcionais aos riscos do sistema e incluir a concepção dos modelos de IA generativa estritamente como ferramentas auxiliares não vinculativas, a eliminação de quaisquer mecanismos que obriguem ou pressionem os juízes a justificar o afastamento de sugestões algorítmicas, a realização de avaliações rigorosas antes de qualquer utilização operacional de ferramentas de IA, a condução de auditorias técnicas independentes e a disponibilização de mecanismos e formação contínua para prevenir o enviesamento por automatização e preservar o juízo humano autónomo³².

2.2.1.6. Transparência e fundamentação adequada

61. Deve ficar claramente documentado quando a IA generativa é utilizada em qualquer aspeto da tomada de decisão judicial³³. As partes têm o direito de saber quais os elementos de uma decisão judicial que foram assistidos por sistemas de IA generativa.

2.2.1.7. Direito de acesso a um tribunal

62. Este direito fundamental pode ser comprometido por sistemas automatizados que substituam ou dificultem o acesso à decisão judicial humana, afetando o direito ao julgamento por um juiz humano.

63. As salvaguardas técnicas devem garantir um acesso efetivo a um juiz humano, limitar a IA generativa a funções estritamente auxiliares e proibir a automatização total das decisões judiciais³⁴.

2.2.2. Direito à liberdade e à segurança (artigo 5.º da CEDH)

64. As decisões que afetem a liberdade pessoal exigem o mais elevado nível de supervisão humana e de fundamentação. Os sistemas de IA generativa que

³¹ Salvi, F., Horta Ribeiro, M., Gallotti, R., & West, R. (2024). On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial. arXiv preprint arXiv:2403.14380

³² A Ferramenta de Avaliação da CEPEJ identifica como risco n.º 12: «Risco de enfraquecimento e limitação da responsabilização do juiz através da utilização de IA não explicável». O quadro da HUDERIA permite uma avaliação recorrente dos potenciais impactos adversos na autonomia judicial.

³³ Comissão Europeia para a Eficácia da Justiça. (31 de janeiro de 2025). Reflexões da AIAB sobre o uso da inteligência artificial nos sistemas judiciais (CEPEJ-AIAB(2025)1Rev5).

³⁴ A Ferramenta de Avaliação da CEPEJ identifica como risco n.º 6: «Risco de a IA substituir o acesso ao juiz».

influenciem decisões de detenção, prisão preventiva ou medidas cautelares sem revisão humana adequada e fundamentação transparente representam riscos significativos para as liberdades fundamentais³⁵.

65. A introdução da IA nas funções judiciais deve seguir uma abordagem progressiva e incremental, inicialmente limitada a tarefas administrativas ou procedimentos normalizados com resultados previsíveis. A exatidão, eficácia e segurança da IA generativa devem ser rigorosamente avaliadas antes da sua expansão para funções judiciais mais sensíveis.

66. As salvaguardas técnicas devem incluir sistemas de IA generativa que promovam alternativas à privação da liberdade, explicações obrigatórias, compreensíveis para humanos, sobre recomendações geradas por IA que afetem a liberdade, e uma revisão humana completa de todas as decisões assistidas por IA em que possa ser aplicada uma medida privativa da liberdade.

2.2.3. Direito à vida privada e à proteção de dados (artigo 8.º da CEDH)

67. A privacidade e a proteção dos dados pessoais devem ser asseguradas quando sejam utilizados sistemas de IA generativa. Toda a tecnologia e todos os dados utilizados devem permanecer sob o controlo exclusivo do poder judicial, a fim de prevenir acessos não autorizados, fugas de informação ou interferências externas e preservar a soberania da instituição sobre a informação. Os dados judiciais devem permanecer sob o controlo exclusivo das autoridades judiciais. Os tribunais devem manter a capacidade de mudar de fornecedor de IA sem perda de funcionalidade, bem como o direito de verificar algoritmos e dados de treino através de processos de auditoria contínua.

68. As salvaguardas técnicas podem incluir a pseudonimização local obrigatória dos dados pessoais antes do seu carregamento no sistema de IA, a execução de modelos de IA em infraestruturas sob controlo governamental ou em infraestruturas comerciais disponíveis — neste último caso, mediante acordos contratuais formais entre autoridades públicas e prestadores de serviços tecnológicos — e a implementação de protocolos rigorosos de proteção de dados³⁶.

2.2.4. Direito a um recurso efetivo (artigo 13.º da CEDH)

³⁵ O artigo 5.º, n.º 1, da CEDH protege o direito à liberdade, sujeito apenas a exceções legais estritamente definidas. O artigo 5.º, n.º 2, estabelece o direito de ser informado «das razões da sua detenção». O artigo 5.º, n.º 4, garante o direito à revisão judicial da legalidade da detenção.

³⁶A Ferramenta de Avaliação da CEPEJ identifica como risco nº 2: “Risco de divulgação de dados pessoais ou segredos comerciais”. A estrutura de análise de impacto na privacidade da HUDERIA aborda especificamente os riscos de divulgação não autorizada de dados confidenciais através da sua metodologia de análise de risco COBRA.

69. Os sistemas automatizados baseados em IA não devem dificultar o acesso efetivo a meios de recurso nem à fiscalização judicial³⁷.

70. As salvaguardas técnicas devem assegurar a revisão humana obrigatória das decisões assistidas por IA e estabelecer procedimentos específicos de reação para impugnar decisões automatizadas ou “influenciadas por IA”.

2.2.5. Proibição da discriminação (artigo 14.º da CEDH e Protocolo n.º 12)

2.2.5.1. Princípio da não discriminação

71. Os enviesamentos algorítmicos nos sistemas de IA generativa representam riscos significativos de reprodução ou amplificação de discriminações existentes relativamente a características protegidas, podendo violar princípios fundamentais de igualdade³⁸.

72. As salvaguardas técnicas devem incluir auditorias periódicas de enviesamento, limpeza e balanceamento de dados, termos de regularização da equidade durante o treino dos modelos (tais como restrições de paridade demográfica ou de probabilidades equalizadas), testes contra factuais/A-B e monitorização contínua de métricas de discriminação³⁹.

2.2.5.2. Neutralidade e prevenção ativa de enviesamentos

73. Os sistemas de IA devem ser avaliados periodicamente para detetar e mitigar enviesamentos, assegurando que não perpetuam discriminações existentes nem criam formas de desigualdade na tomada de decisão judicial. A monitorização sistemática de enviesamentos relativamente a todas as características protegidas é obrigatória.

2.2.6. Salvaguardas adicionais

2.2.6.1. Explicabilidade e fundamentação transparente

³⁷ O artigo 13.º da CEDH estabelece que «Toda a pessoa cujos direitos e liberdades consagrados na presente Convenção forem violados terá direito a um recurso efetivo perante uma autoridade nacional, mesmo que a violação tenha sido cometida por pessoas que agissem no exercício de funções oficiais».

³⁸ O artigo 14.º da CEDH estabelece que «O gozo dos direitos e liberdades enunciados na presente Convenção será assegurado sem discriminação por qualquer motivo, como sexo, raça, cor, língua, religião, opinião política ou outra, origem nacional ou social, associação a uma minoria nacional, propriedade, nascimento ou outro estatuto».

³⁹ A Ferramenta de Avaliação da CEPEJ identifica como risco nº 10: «Risco de discriminação ou amplificação da discriminação». O processo de envolvimento das partes interessadas da HUDERIA permite identificar impactos diferenciados em grupos vulneráveis.

74. Os sistemas de IA utilizados na esfera judicial devem fornecer explicações claras e compreensíveis do percurso seguido para cada resultado ou recomendação, permitindo que os operadores judiciais os verifiquem criticamente. Para esse efeito, os sistemas devem divulgar o seu processo interno de inferência de forma transparente e auditável, assegurando uma conformidade substancial e evitando simulações de cumprimento meramente superficiais que ocultem a sua lógica real⁴⁰. Os sistemas devem ser capazes de indicar as fontes específicas de cada afirmação gerada. Apenas devem ser utilizados modelos interpretáveis que revelem os seus processos internos de tomada de decisão.

2.2.6.2. Proporcionalidade e cautela na implementação da IA generativa

75. A introdução da IA nas funções judiciais deve seguir uma abordagem progressiva e incremental, inicialmente limitada a tarefas administrativas ou a procedimentos normalizados com resultados previsíveis. A exatidão, a eficácia e a segurança da IA generativa devem ser rigorosamente avaliadas antes da sua extensão a funções judiciais mais sensíveis.

2.2.6.3. Responsabilidade do Estado por danos causados pela utilização de IA generativa

76. O Estado deve assumir a responsabilidade por quaisquer danos causados por erros judiciais ou falhas dos sistemas de IA generativa, em conformidade com os princípios existentes da responsabilidade do Estado. Nem a autonomia nem a complexidade tecnológica da IA generativa devem ser invocadas para isentar o Estado do seu dever de reparação⁴¹.

2.2.6.4. Princípio da equidade processual e da eficiência

77. A introdução da IA na administração da justiça não deve resultar numa deterioração das garantias processuais nem num prejuízo da tutela jurisdicional efetiva; pelo contrário, deve reforçar os direitos processuais e a eficácia judicial. Por exemplo, deve melhorar o acesso ao tribunal, facilitar o acesso à prova ou melhorar os meios de apresentação das alegações das partes. Além disso, a implementação da IA generativa deve traduzir-se em melhorias mensuráveis no tempo de resposta, na clareza da comunicação, na consistência das decisões e na eficiência da gestão processual. Os tribunais devem manter a capacidade de regressar a sistemas anteriores caso a IA generativa não melhore objetivamente

⁴⁰ Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Marks, S., Treutlein, J., Belonax, T., Chen, J., Duvenaud, D., Khan, A., Michael, J., Mindermann, S., Perez, E., Petrini, L., Uesato, J., Kaplan, J., Shlegeris, B., Bowman, S. R., & Hubinger, E. (2024). Falsificação de alinhamento em grandes modelos de linguagem [arXiv preprint arXiv:2412.14093v2]. <https://doi.org/10.48550/arXiv.2412.14093>

⁴¹ Os indicadores de avaliação devem incluir inquéritos de satisfação dos utilizadores, comparação de métricas de desempenho com sistemas pré-IA, feedback das partes interessadas, incluindo juízes, funcionários judiciais, advogados e cidadãos, e avaliação das melhorias nos prazos processuais.

os serviços judiciais. Deverão ser realizadas avaliações regulares para verificar o cumprimento deste princípio.

3. GLOSSÁRIO

Princípio da neutralidade tecnológica

É geralmente definido como a liberdade de indivíduos e organizações escolherem a tecnologia mais apropriada e adequada às suas necessidades e requisitos para desenvolvimento, aquisição, utilização ou comercialização, sem dependências do conhecimento envolvido enquanto informação ou dados⁴².

Regularizadores de equidade durante o treino dos modelos

Técnicas matemáticas específicas, tais como:

- Regularizadores de paridade demográfica: penalizam o modelo quando as taxas de previsão diferem significativamente entre grupos demográficos
- Regularizadores de probabilidades equalizadas: asseguram taxas semelhantes de verdadeiros positivos e falsos positivos entre grupos protegidos
- Restrições de equidade individual: garantem que indivíduos semelhantes recebem tratamento semelhante
- Remoção de viés: utilização de redes adversárias para remover informação relativa a características protegidas das representações aprendidas

⁴² Ver: https://icannwiki.org/Technology_neutrality