

DIREITOS AUTORAIS & MACHINE LEARNING: A LIBERDADE DE PROSPEÇÃO DE TEXTOS E DADOS¹

Copyright and Machine Learning: The Freedom of Text and Data Mining

Alexandre Libório Dias Pereira²

RESUMO:

O desenvolvimento da Inteligência Artificial justificou a introdução de uma nova exceção relativamente à prospeção de textos e dados para fins de investigação científica por parte de organismos de investigação e entidades responsáveis pelo património cultural. Esta exceção é imperativa, não admitindo cláusula em contrário nem proteção técnica que a impeça. Além desta exceção de interesse geral, é ainda prevista uma exceção de prospeção de textos e dados para outros fins, de natureza supletiva, e que se aplica também aos programas de computador. É discutível se o novo regime alarga ou, pelo contrário, se restringe a liberdade de inovação no domínio da Inteligência Artificial, não sendo certo de que modo o novo regime de direitos de autor na União Europeia, guiado pelo princípio do elevado nível de proteção da propriedade intelectual, contribuirá para o desenvolvimento da Inteligência Artificial baseado em “aprendizagem da máquina”.

Palavras-chave: Direitos de autor; prospeção de textos e dados; Inteligência Artificial; investigação científica; inovação tecnológica.

ABSTRACT:

The development of Artificial Intelligence justified the introduction of a new exception regarding text and data mining for purposes of scientific research by research organizations and entities responsible for cultural heritage. This exception is imperative, not admitting clauses to the contrary nor technical protection that hinders it. In addition to this exception of general interest, an exception is also foreseen for text and data mining for other purposes, which applies by default, and also to computer programs. It is doubtful whether the new regime extends or, on the contrary, restricts the freedom of innovation in the field of Artificial Intelligence, as it is far from clear how the new EU copyright rules, guided by the principle of a high level of intellectual property protection, will contribute to the development of Artificial Intelligence based upon Machine Learning.

Keywords: *Copyright; text and data mining; Artificial Intelligence; scientific research; technological innovation.*

¹ Texto de apoio à palestra ministrada na mesa sobre Inteligência Artificial e Novas Fronteiras da Propriedade Intelectual do XVII Congresso de Direito de Autor e Interesse Público organizado pelo Grupo de Estudos de Direito Autoral e Industrial (GEDAI) da Universidade Federal do Paraná (UFPR) e que teve lugar em Curitiba, nos dias 6, 7 e 8 de novembro de 2023.

² Universidade de Coimbra, Professor Associado da Faculdade de Direito e Investigador do Instituto Jurídico.

SUMÁRIO

INTRODUÇÃO: UMA LEI DE DIREITOS DE AUTOR PARA A PROMOÇÃO DA APRENDIZAGEM DA MÁQUINA? **1. PROSPEÇÃO DE TEXTOS E DADOS PARA FINS DE INVESTIGAÇÃO CIENTÍFICA POR ORGANISMOS DE INVESTIGAÇÃO E INSTITUIÇÕES DO PATRIMÔNIO CULTURAL;** **2. PROSPEÇÃO DE TEXTOS E DADOS POR OUTRAS ENTIDADES E/OU PARA OUTROS FINS;** **3. A TRANSPOSIÇÃO EFETUADA PELO DECRETO-LEI 47/2023. CONCLUSÃO.**

INTRODUÇÃO: UMA LEI DE DIREITOS DE AUTOR PARA A PROMOÇÃO DA APRENDIZAGEM DA MÁQUINA?

A Inteligência Artificial (doravante, IA) foi identificada pela Comissão Europeia como um setor crítico para a Europa, considerando que o “aumento da capacidade computacional, a disponibilidade de dados e a evolução dos algoritmos fizeram da IA uma das tecnologias mais estratégicas do século XXI”³. Todavia, o desenvolvimento da IA enfrenta constrangimentos legais e incerteza jurídica, nomeadamente no que respeita à utilização pelos sistemas de IA de obras literárias e artísticas protegidas por direitos de autor, o mesmo valendo para os direitos conexos, os programas de computador e as bases de dados.

Do ponto de vista meramente técnico, os sistemas de IA reproduzem obras e outros conteúdos no processo de elaboração de novos conteúdos, sejam textos, sons e/ou imagens fixas ou animadas, e até software. No domínio das bases de dados, os sistemas de IA procedem à extração de dados da base. Estes atos de reprodução ou extração estão sujeitos a direitos de autor ou direitos especiais?

Ainda em fase de trabalhos preparatórios sobre a Diretiva 2019/790 sobre direitos de autor no mercado único digital⁴, a Comissão Europeia considerou muito útil para o desenvolvimento da IA a nova exceção ou

³ **Inteligência artificial para a Europa**, COM(2018) 237 final, Bruxelas, 25.4.2018, p. 3.

⁴ Diretiva (UE) 2019/790 do Parlamento Europeu e do Conselho de 17 de abril de 2019 relativa aos direitos de autor e direitos conexos no mercado único digital e que altera as Diretivas 96/9/CE e 2001/29/CE.

limitação obrigatória para a utilização de tecnologias de prospeção de textos e dados no domínio da investigação científica⁵.

Mais tarde, já depois de aprovada a Diretiva, o Parlamento Europeu adotou uma resolução sobre direitos de propriedade intelectual para o desenvolvimento das tecnologias de inteligência artificial⁶, nos termos da qual será possível a proteção de obras geradas autonomamente por IA com “recurso a material protegido por direitos de autor, apenas se o titular dos direitos de autor tiver concedido autorização, a menos que se apliquem exceções ou limitações aos direitos de autor” (para. 15), em especial “a exceção relativa à prospeção de textos e dados prevista na diretiva relativa aos direitos de autor e direitos conexos no mercado único digital” (para. 17), recordando ainda “o dever ético da União de apoiar o desenvolvimento em todo o mundo, facilitando a cooperação transfronteiras em matéria de IA, nomeadamente através de limitações e exceções no que respeita à investigação transfronteiras e à pesquisa de textos e dados, conforme estabelece a Diretiva relativa aos direitos de autor e direitos conexos no mercado único digital”⁷.

⁵ Inteligência artificial para a Europa, *cit.*, p. 11. Sobre os trabalhos preparatórios sobre o tema, *vd.* Jean-Paul Triaille/Jérôme de Meeûs d’Argenteuil/Amélie de Francquenm, **Study on the legal framework of text and data mining (TDM)**, 2014; Christophe Geiger/Giancarlo Frosio/Oleksandr Bulayenko, «The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects», European Parliament, 2018.

⁶ Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre os direitos de propriedade intelectual para o desenvolvimento de tecnologias ligadas à inteligência artificial, para. 20. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_PT.html. Acesso: 20 out. 2023.

Esta resolução é uma entre várias resoluções do Parlamento Europeu sobre IA e surge em linha com a agenda da Comissão Europeia anunciada no seu Livro Branco sobre a inteligência artificial – Uma abordagem europeia virada para a excelência e a confiança, COM(2020) 65 final, Bruxelas, 19.2.2020. Mais recentemente, *vd.* Resolução do Parlamento Europeu, de 3 de maio de 2022, sobre a inteligência artificial na era digital, disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_PT.html. Acesso em 20 out. 2023.

⁷ A política de desenvolvimento da IA não poderá fazer tábua rasa dos direitos de autor, nem isentar os utilizadores de sistemas de IA relativamente a violações aos direitos de autor, atribuindo a responsabilidade ao Robot. Como referem MARTIN SENFLEBEN/LAURENS BUIJTELAAR, «Robot Creativity: An Incentive-Based Neighbouring Rights Approach», **European Intellectual Property Review**, Vol. 42/12 (2020), p. 797-812,

Assim, é na Diretiva 2019/790 que devemos procurar respostas para a questão de saber se e como servem atualmente os direitos de autor o desenvolvimento da IA, nomeadamente na chamada aprendizagem da máquina (*machine learning*). A primeira lei de direitos de autor⁸ intitulava-se justamente “An Act for the Encouragement of Learning”, ou seja, uma lei para a promoção da aprendizagem. Será este o tempo do copyright para a promoção da aprendizagem da máquina?

1 PROSPEÇÃO DE TEXTOS E DADOS PARA FINS DE INVESTIGAÇÃO CIENTÍFICA POR ORGANISMOS DE INVESTIGAÇÃO E INSTITUIÇÕES DO PATRIMÓNIO CULTURAL

A Diretiva 2019/790 introduz exceções ou limitações “obrigatórias” para usos de tecnologias de prospeção de texto e dados, ilustração para ensino no ambiente digital e para preservação do património cultural (cons. 5). Estas exceções somam-se às já previstas designadamente nas Diretivas 96/9, 2001/29 e 2009/24, mas agora são obrigatórias e não meramente facultativas. Considera-se que a nova exceção de prospeção de textos e dados desempenha um papel importante para o desenvolvimento da IA.

Para começar, a prospeção de textos e dados é considerada um novo tipo de utilização tornado possível pelas tecnologias digitais e ainda não previsto na legislação da UE (cons. 8). É a chamada análise computacio-

812: “Instead of seeing the grant of protection as a stimulus for stronger efforts to develop the full potential of creative AI machines, robot users may eschew the right holder status to escape liability for potential infringements”. Sobre esta questão, os deputados ao Parlamento Europeu (MEP) consideraram que a IA generativa (por exemplo, GPT) deve divulgar a fonte do conteúdo gerado, em particular concebendo o sistema por defeito para prevenir conteúdos ilegais. Vd. DRAFT Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (16/5/2023), disponíveis em <https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf>. Acesso em 15 set. 2023.

⁸ THE BRITISH COPYRIGHT STATUTE OF ANNE (1710). Disponível em: <https://www.copyright-history.com/anne.html>. Acesso em 18 ago. 2023.

nal automatizada de informações em formato digital, como textos, sons, imagens ou dados. A prospeção de textos e dados é uma ferramenta muito valiosa para a investigação e a inovação, uma vez que processa grandes quantidades de informação com o objetivo de obter novos conhecimentos e descobrir novas tendências. Por prospeção de textos e dados entende-se “qualquer técnica de análise automática destinada à análise de textos e dados em formato digital, a fim de produzir informações, tais como padrões, tendências e correlações, entre outros”⁹.

Para superar a insegurança jurídica sobre o relevo jurídico-autoral da prospeção de textos e dados, a Diretiva estabelece que as universidades e outros organismos de investigação, bem como as instituições responsáveis pelo património cultural podem fazer prospeção de textos e de dados sem autorização dos titulares de direitos, por ex. quando os dados são normalizados nesse processo, mesmo que esses atos sejam protegidos por direitos autorais, direitos conexos, direitos especiais, como o direito *sui generis* do produtor de base de dados ou o novo direito do editor de publicações de imprensa no que diz respeito a utilizações em linha.¹⁰

⁹ Art. 2.º, n.º 2, da Diretiva 2019/790.

¹⁰ Os dados pessoais também suscitam preocupações sobre as utilizações de IA, que podem ser permitidas para fins de investigação científica de interesse geral ao abrigo do Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados, e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados), JO L 119 de 4.5.2016, p. 1 – vd. artigos 9.º/2-j) (tratamento de categorias especiais de dados pessoais), 14.º/5-b) (informações a fornecer quando os dados pessoais não tenham sido obtidos junto do titular dos dados), 17.º/3-d) (direito ao apagamento - «direito a ser esquecido»), 21.º/6 (direito de oposição) e, em especial, 89.º (salvaguardas e derrogações relativas ao tratamento para fins de arquivo de interesse público, para fins de investigação científica ou histórica ou para fins estatísticos); vd. também os considerandos 26, 33, 50 a 53, 62, 65, 113, 156 e 157 a 160, e o documento do PARLAMENTO EUROPEU, **How the General Data Protection Regulation changes the rules for scientific research** (EPRS Brussels 2019). No que diz respeito aos dados não pessoais, o princípio é que devem estar disponíveis gratuitamente em formatos abertos e legíveis por máquina para alimentar a aprendizagem automática e a IA. No considerando 9 do Regulamento (UE) 2018/1807 do Parlamento Europeu e do Conselho, de 14 de novembro de 2018, relativo a um quadro para o livre fluxo de dados não pessoais na União Europeia, JO L 303 de 28.11.2018, p. 59, lê-se que “A internet das coisas, a inteligência artificial e a aprendizagem automática, que estão em expansão, representam grandes fontes de dados não pessoais, por exemplo, em consequência da sua utilização em processos automa-

Esta exceção de prospeção de textos ou dados para fins de investigação científica aproveita apenas a organismos de investigação e instituições responsáveis pelo património cultural. Por “organismo de investigação” entende-se “uma universidade, incluindo as suas bibliotecas, um instituto de investigação ou qualquer outra entidade cujo principal objetivo seja a realização de investigação científica ou o exercício de atividades didáticas que envolvam igualmente a realização de investigação científica, sem fins lucrativos ou para reinvestir a totalidade dos lucros na investigação científica ou no quadro de uma missão de interesse público reconhecida por um Estado-Membro, de modo que o acesso aos resultados provenientes dessa investigação científica não possa beneficiar em condições preferenciais uma empresa que exerça uma influência decisiva sobre esse organismo”. Já por “instituição responsável pelo património cultural”, entende-se “uma biblioteca ou um museu acessíveis ao público, um arquivo ou uma instituição responsável pelo património cinematográfico ou sonoro”¹¹.

A exceção de prospeção de textos ou dados para fins de investigação científica prevalece sobre eventuais condições de assinatura de publicações ou de termos de licenças de acesso aberto que excluam a prospeção de textos e de dados (cons. 10). A exceção é limitada ao direito exclusivo de reprodução e ao direito de impedir a extração de base de dados, e é prevista a favor apenas de universidades e outros organismos de investigação, e instituições responsáveis pelo património cultural, embora seja extensível aos seus parceiros privados em atividades de investigação no âmbito de parcerias público-privadas (cons. 11). Todavia, esta exceção depende de os respetivos beneficiários terem “acesso legal” aos conte-

tizados de produção industrial. Exemplos concretos de dados não pessoais incluem conjuntos de dados agregados e anonimizados utilizados para a análise de grandes volumes de dados, os dados relativos à agricultura de precisão que podem ajudar a controlar e a otimizar a utilização de pesticidas e de água ou ainda dados sobre as necessidades de manutenção de máquinas industriais”. O mesmo princípio é válido relativamente à reutilização de informação do setor público, nos termos da Diretiva (UE) 2019/1024 do Parlamento Europeu e do Conselho, de 20 de junho de 2019, relativa aos dados abertos e à reutilização de informações do setor público, OJ L 172, 26.6.2019, p. 56–83.

¹¹ Art. 1.º, n.ºs 1 e 3, da Diretiva 2019/790, respetivamente.

údos protegidos, podendo tal acesso basear-se numa política de acesso aberto ou em acordos entre titulares de direitos e organismos de pesquisa ou instituições de património cultural (por ex., assinaturas), ou por outros meios legais, como, por ex., conteúdos livremente disponíveis em linha (cons. 14).

Assim, não carecem de autorização do titular de direitos as “reproduções e extrações efetuadas por organismos de investigação e por instituições responsáveis pelo património cultural para a realização de prospeção de textos e dados de obras ou outro material protegido a que tenham acesso legal para efeitos de investigação científica”¹².

A exceção de prospeção de textos de dados justifica o armazenamento de cópias para fins de investigação científica, incluindo para verificação subsequente dos resultados da investigação¹³, desde que o armazenamento seja feito em ambiente seguro. Fica a cada Estado-Membro decidir, após debate com as partes interessadas relevantes, sobre outras disposições específicas para a conservação das cópias, incluindo a possibilidade de nomear organismos de confiança para o efeito. Todavia, as condições de segurança não devem restringir indevidamente a aplicação da exceção, antes devendo ser proporcionadas e limitadas ao necessário para

¹² Art. 3.º, n.º 1, da Diretiva 2019/790,

¹³ Um direito acessório de digitalizar obras por bibliotecas de acesso público foi reconhecido pelo TJUE no acórdão *Eugen Ulmer*, C-117/13, ECLI:EU:C:2014:2196, para. 43, o qual adotou a doutrina dos direitos dos utilizadores, posteriormente confirmada no acórdão *Funke Medien* de 29 de julho de 2019, C-469/17, ECLI:EU:C:2019:623, para. 70 (“estas próprias exceções ou limitações comportam direitos em benefício dos utilizadores de obras ou outros materiais protegidos”), e no acórdão *Polónia c. Parlamento e Conselho*, de 26 de abril de 2022, C-401/19, ECLI:EU:C:2022:297, para. 88 (“garantir os direitos dos utilizadores”), e tal como defendera o Advogado-Geral HENRIK SAUGMANDSGAARD ØE nas *Conclusões* apresentadas em 15 de julho de 2021, ECLI:EU:C:2021:613 (“o legislador da União reconheceu expressamente direitos subjetivos em matéria de direitos de autor aos utilizadores dos serviços de partilha”). Defendendo a doutrina «users’ rights» no copyright norte-americano foi defendida por L. Ray Patterson e Stanley W. Lindberg, **The Nature of Copyright: A Law of Users’ Rights**, The University of Georgia Press Athens, 1991; em Portugal, Alexandre L. Dias Pereira, **Direitos de Autor e Liberdade de Informação**, Almedina, Coimbra, 2008, p. 649, Tito Rendas, “Are copyright-permitted uses ‘exceptions’, ‘limitations’ or ‘user rights’? The special case of Article 17 CDSM Directive”, **Journal of Intellectual Property Law & Practice** vol. 17 (2022), p. 54–64.

conservar as cópias de forma segura e impedir a utilização não autorizada. De igual modo, as condições de segurança não devem impedir outras utilizações para fins de investigação científica, como a revisão científica por pares e a investigação conjunta, que se consideram já abrangidas pela exceção ou limitação prevista no art. 5.º, n.º 3, alínea a), da Diretiva 2001/29/CE (cons. 5). De resto, esta exceção não prejudica as restantes exceções ou limitações aos direitos exclusivos. Deste modo, como se lê no preâmbulo, “(9) A prospeção de textos e dados pode ser igualmente realizada em relação aos factos em si ou aos dados que não estão protegidos por direitos de autor e, nesses casos, não é necessária qualquer autorização ao abrigo do direito em matéria de direitos de autor”, assim como não envolver atos de reprodução ou envolver reproduções abrangidas pela exceção obrigatória sobre os atos de reprodução temporária prevista no art. 5.º, n.º 1, da Diretiva 2001/29/CE, “que deverá continuar a ser aplicada às técnicas de prospeção de textos e dados que não impliquem fazer cópias dos materiais para além do âmbito desta exceção” (cons. 9). Ou seja, o princípio é o de que a exceção de prospeção de textos e dados não deve ser interpretada *a contrario sensu* para alargar o objeto e o conteúdo do direito exclusivo, nem apertar o alcance das exceções e limitações já existentes.

Assim, as cópias de obras ou de outro material protegido efetuadas no âmbito da prospeção de textos e dados devem ser armazenadas com um nível de segurança adequado e podem ser conservadas para fins de investigação científica, incluindo para a verificação dos resultados da investigação, podendo os titulares de direitos aplicar medidas para assegurar a segurança e a integridade das redes e bases de dados em que as obras ou outro material protegido são acolhidos na medida do que for necessário para alcançar esse objetivo. Para o efeito, os Estados-Membros incentivarão os titulares de direitos, os organismos de investigação e as instituições responsáveis pelo património cultural a definir melhores práticas previamente acordadas¹⁴.

¹⁴ Art. 3.º, n.º 2, 3 e 4 da Diretiva 2019/790. O regime jurídico da segurança informática é estabelecido pela Lei n.º 46/2018, de 13 de agosto, que transpõe a Diretiva (UE) 2016/1148, do Parlamento Europeu e do Conselho, de 6 de julho de 2016, relativa a medidas destinadas a garantir um elevado nível comum de segurança das redes e dos sistemas de informação em toda a União Europeia, já revogada e substituída pela Di-

Será devida uma compensação equitativa pelas reproduções efetuadas pela prospeção de textos e dados para fins de investigação científica? Nos termos do preâmbulo, não se justifica compensar os titulares de direitos pelas utilizações permitidas pela nova exceção, uma vez que o dano potencial que sofrem é mínimo, tendo em conta que é limitada a entidades que realizam investigação científica (cons. 17).

2 PROSPEÇÃO DE TEXTOS E DADOS POR OUTRAS ENTIDADES E/OU PARA OUTROS FINS

Ao lado da exceção para fins de investigação científica por organismos de investigação e instituições do património cultural é prevista uma outra exceção de prospeção de textos e dados por outras entidades e/ou para outros fins. Como se lê no preâmbulo: “Para além da sua importância no contexto da investigação científica, as técnicas de prospeção de textos e dados são amplamente utilizadas, tanto pelas entidades públicas, como privadas para analisar grandes quantidades de dados em diferentes domínios da vida e para vários fins, nomeadamente para serviços públicos, decisões empresariais complexas e para o desenvolvimento de novas aplicações ou tecnologias.”¹⁵

Entende-se, todavia, que a prospeção de dados para estes outros fins e por outras entidades deve poder ser objeto de licenças por parte dos titulares de direitos, embora supletivamente não seja necessária autorização. Considera-se existir insegurança jurídica relativamente às reproduções e extrações efetuadas para efeitos de prospeção de textos e dados mesmo quando o acesso ao conteúdo seja legal, em especial no que respeita às condições da exceção de reprodução temporária prevista no art. 5.º, n.º 1, da Diretiva 2001/29/CE.

Assim, a solução encontrada e que visa igualmente, além de proporcionar maior segurança jurídica, “incentivar a inovação também no

retiva (UE) 2022/2055 do Parlamento Europeu e do Conselho, de 14 de dezembro de 2022.

¹⁵ Considerando (18) da Diretiva 2019/790.

setor privado” (cons. 18) foi estabelecer uma exceção geral de reproduções e extrações de obras ou outro material protegido para fins de prospeção de textos e dados e de armazenamento de cópias durante o tempo necessário para esses fins. Esta exceção pressupõe também que o acesso ao conteúdo seja legal, “nomeadamente sempre que tenha sido disponibilizado em linha ao público e desde que os titulares dos direitos não tenham reservado, de forma adequada, o direito de efetuar reproduções e extrações para prospeção de textos e dados”¹⁶.

Os titulares de direitos podem reservar o direito de prospeção de dados mediante a utilização de meios de leitura ótica, incluindo metadados e condições gerais de um sítio Internet ou de um serviço, relativamente a conteúdos que tenham sido disponibilizados em linha ao público, ou por acordos contratuais ou por uma declaração unilateral nos restantes casos. Em qualquer caso, a exceção supletiva de prospeção de textos e dados não deve prejudicar a exceção prevista obrigatoriamente para fins de investigação científica, nem a exceção de reprodução temporária prevista no art. 5.º, n.º 1, da Diretiva 2001/29/CE.

Em todo o caso, é necessário não esquecer que as reproduções realizadas no âmbito das atividades de prospeção de textos e dados para fins diferentes da investigação científica e/ou por outras entidades podem ter ‘valor económico distinto por si só’ e, portanto, não serem completamente estranhas aos direitos de autor, tendo em conta a amplitude das definições dos direitos de reprodução e de extração¹⁷.

Com efeito, não é fácil traçar a linha entre as reproduções transitórias ou incidentais, incluindo atos de browsing e de armazenamento

¹⁶ Considerando (18), para. 2, da Diretiva 2019/790.

¹⁷ No acórdão *Infopaq I*, o TJUE concluiu que um ato que ocorre durante um processo de recolha de dados, que consiste em armazenar um extrato de uma obra protegida composto por 11 palavras e imprimir esse extrato, sem destruir, em si mesmo, esse suporte, é suscetível de ser abrangido pelo conceito de reprodução parcial na aceção do artigo 2.º da Diretiva 2001/29: acórdão de 16 de julho de 2009, C-5/08, ECLI:EU:C:2009:465, paras. 53, 69-7. Além disso, no processo *British Horseracing*, o TJUE considerou que a Diretiva Base de Dados fornecia uma “definição ampla” do direito de extração, a fim de proteger o produtor contra atos suscetíveis de o privarem de receitas necessárias para financiar o investimento - acórdão de 9 de novembro de 2004, C-203/02, ECLI:EU:C:2004:695, n.º 51.

temporário (*caching*)¹⁸, e o que agora é abrangido pelas atividades de prospeção de textos e dados, como sucede com a coleta de dados de treinamento de IA na web e até que ponto será necessária uma nova exceção para efeito¹⁹.

A licitude das reproduções transitórias ou incidentais, nos termos do art. 5.º/1 da Diretiva 2001/29 é fundamental para permitir a transmissão eficiente numa rede entre terceiros por um intermediário, ou a utilização lícita de uma obra ou outro material, sendo que, em particular, os atos de reprodução em causa não devem ter “valor económico em si”.

Uma vez que a prospeção de textos e dados implica reproduções que, só por si, têm valor económico então não são excluídas do direito de reprodução. Os direitos de autor expandem-se em linha com a expansão do valor económico da atividade de prospeção de textos e dados e, portanto, ficam sujeitos à autorização do titular do direito ou à restrição legal do direito exclusivo.

Os titulares de direitos podem impedir a prospeção de textos e dados para outros fins que não a investigação científica. Apesar de supletiva, esta exceção geral de prospeção de textos e dados aplica-se igualmente aos programas de computador, cuja proteção jurídica já previa a liberdade de análise de programas como objeto de pesquisa científica ou de ensino, assim como a descompilação das interfaces do programa para fins de interoperabilidade com outros programas²⁰. Não obstante se sustentar que a prospeção de textos e dados deveria igualmente aplicar-se a programas de computador quando efetuada para fins de investigação científica²¹, a

¹⁸ Diretiva 2001/29, considerando (33).

¹⁹ Jan Bernd Nordemann, Jonathan Pukas, «Copyright exceptions for AI training data—will there be an international level playing field?», *JIPLP* 17/12 (2022), p. 973-4 (“time will tell whether perhaps a new exception provision under copyright law is needed for the web harvesting of AI training data in Europe”).

²⁰ Art. 15.º, n.º 3, e art. 6.º, respetivamente, da Diretiva n.º 91/250/CEE, do Conselho, de 14 de maio, relativa ao regime de proteção jurídica dos programas de computador, entretanto revogada e substituída pela Diretiva 2009/24/CE do Parlamento Europeu e do Conselho de 23 de abril de 2009.

²¹ J. Griffiths/T. Synodinou/R. Xalabarder, **Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market**, Europe-

verdade é que, quando efetuada para outros fins ou até por outras entidades que prossigam fins de investigação científica, mas não se enquadrem na noção de organismos de investigação, a análise de software mediante prospeção de textos e dados é agora apenas supletiva, já não imperativa. Com efeito, é estabelecida uma exceção aos direitos previstos no art. 5.º, alínea a), e 7.º, n.º 1, da Diretiva 96/9 (bases de dados), no art. 2.º Diretiva 2001/29 (direito de reprodução), no art. 4.º, n.º 1, alíneas a) e b), da Diretiva 2009/24 (programas de computador) e no art. 15.º, n.º 1, da diretiva mercado digital (direito do editor de publicações de imprensa), para as reproduções e as extrações de obras e de outro material protegido legalmente acessíveis para fins de prospeção de textos e dados, assim como para a conservação desses conteúdos enquanto for necessário para esses fins, ressalvando-se, todavia, que os titulares de direitos podem reservar expressamente essa utilização de forma adequada, em particular por meio de leitura ótica no caso de conteúdos disponibilizados ao público em linha, sem prejuízo da exceção de prospeção de textos e dados para fins de investigação científica.²²

3 A TRANSPOSIÇÃO EFETUADA PELO DECRETO-LEI 47/2023

A Lei n.º 11/2023, de 22 de março, autorizou o Governo a legislar em matéria de direito de autor e direitos conexos no mercado único digital, transpondo a Diretiva 2019/790. Na matéria em análise limita-se a permitir ao governo “criar exceções e limitações ao exercício do direito de autor e dos direitos conexos, incluindo os direitos sobre os programas de computador e sobre as bases de dados, nos termos dos artigos 3.º, 4.º, 5.º, 6.º, 7.º, 8.º, 9.º, 10.º, 17.º e 24.º da Diretiva (UE) 2019/790” (art. 2.º, alínea b)).

an Copyright Society, 2022: “While it is possible that TDM for research on computer programs can be performed on the grounds of Art 5 (3) of the Computer Programs Directive (the so called ‘black box’ analysis), that provision does not cover acts of translation and adaptation. [...] National lawmakers are encouraged to eliminate the disparities regarding the scope of application of the provisions of Art 3 and 4 DSMD and specifically to provide that TDM for research purposes on computer programs is also allowed.” – p. 16.

²² Art. 4.º, n.º 1, da Diretiva 2019/790.

A Diretiva 2019/790 foi transposta para o direito interno pelo Decreto-Lei n.º 47/2023, de 19 de junho. No que respeita às exceções de prospeção de textos e dados, a transposição consistiu em introduzir duas novas alíneas, a v) e a w), no n.º 2 do artigo 75.º, e no aditamento de um novo número, o 6, ao referido artigo, para prever as noções de “organismo de investigação” e de “prospeção de textos e dados”. Além disso, alterou o artigo 76.º sobre requisitos da utilização livre, introduzindo novos números, designadamente o quatro (sobre armazenamento seguro das reproduções efetuadas para fins de prospeção) e o cinco (sobre medidas de segurança e de integridade das redes e bases de dados onde são conservados os textos e dados objeto de prospeção). A este propósito, o legislador não tirou partido da possibilidade de nomear organismos de confiança para o efeito das medidas de segurança informática e de integridade das redes.

Por outro lado, o Decreto-Lei 43/2023 alterou igualmente o Decreto-Lei 122/2000, de 4 de julho, que transpôs a Diretiva 96/9. Aditou uma nova alínea ao número 1 do artigo 10.º, sob epígrafe exceções, permitindo atos de reprodução e extração de obras e de outro material protegido legalmente acessíveis para fins de prospeção de textos e dados, desde que essa utilização não tenha sido expressamente reservada pelos respetivos titulares de direitos de forma adequada (exceção supletiva de prospeção de textos e dados para outros fins). Além disso, aditou ao art. 15.º duas outras utilizações livres, a saber os “atos de reprodução e extração efetuados por organismos de investigação e por instituições responsáveis pelo património cultural para a realização de prospeção de textos e dados de obras ou outro material protegido a que tenham acesso legal para efeitos de investigação científica”, por um lado (al. e) e os “atos de reprodução e extração de obras e de outro material protegido legalmente acessíveis para fins de prospeção de textos e dados, desde que essa utilização não tenha sido expressamente reservada pelos respetivos titulares de direitos de forma adequada, em particular por meio de leitura ótica no caso de conteúdos disponibilizados ao público em linha”, por outro (al. f).

De notar ainda que o Decreto-Lei 47/2023 não alterou o Decreto-Lei 252/94 sobre proteção jurídica dos programas de computador. A Diretiva 2019/790 estabelece expressamente que “Os Estados-Membros devem

prever uma exceção ou uma limitação aos direitos previstos [...] no artigo 4.º, n.º 1, alíneas a) e b), da Diretiva 2009/24/CE [...], para as reproduções e as extrações de obras e de outro material protegido legalmente acessíveis para fins de prospeção de textos e dados” para fins diferentes da investigação científica ou por entidades que não sejam organismos de investigação ou instituições responsáveis pelo património cultural.

A transposição da Diretiva foi omissa quanto a este ponto. Terá o legislador considerado, possivelmente, que a não previsão da exceção de prospeção de textos e dados para fins de investigação científica no domínio dos programas de computador seria interpretada, *a contrario*, no sentido de esses atos não serem permitidos. Todavia, parece-nos que essa utilização livre já resultaria da liberdade de análise de programas como objeto de pesquisa científica ou de ensino, assim como a descompilação das interfaces do programa para fins de interoperabilidade com outros programas²³. Em qualquer caso, a omissão do legislador é criticável, embora se reconheça que a natureza supletiva da exceção de prospeção de textos e dados para fins diferentes da investigação científica contribuirá para limitar ainda mais a inovação tecnológica concorrencial no *software*, que é um setor crítico da IA.

CONCLUSÃO

O desenvolvimento da IA justificou a introdução de uma exceção relativamente à prospeção de textos e dados para fins de investigação científica por parte de organismos de investigação e entidades responsáveis pelo património cultural. Esta exceção é obrigatória e imperativa, não admitindo cláusula em contrário nem proteção técnica que a impeça.

Além desta exceção de interesse geral, é ainda prevista uma exceção de prospeção de textos e dados para outros fins, mas com algumas

²³ Art. 10.º, n.º 2, e art. 7.º, respetivamente, do Decreto-Lei n.º 252/94, de 20 de outubro, que transpõe para a ordem jurídica interna a Diretiva n.º 91/250/CEE, do Conselho, de 14 de maio, relativa ao regime de proteção jurídica dos programas de computador, entretanto revogada e substituída pela Diretiva 2009/24/CE do Parlamento Europeu e do Conselho de 23 de abril de 2009.

nuances. Primeiro, apesar de ser também obrigatória, já é apenas supletiva, podendo o titular de direitos reservar esta nova utilização. Segundo, aplica-se igualmente aos programas de computador, acrescentando à liberdade de descompilação para fins de interoperabilidade e à liberdade de análise de programas como objeto de pesquisa científica ou de ensino.

O desenvolvimento da IA, no que respeita à utilização em prospeção de textos e dados, introduz para os titulares de direitos o ónus de reservar expressamente esta nova utilização, que de outro modo é supletivamente autorizada por lei, e até imperativa, se realizada para fins de investigação científica.

A razão de ser desta nova exceção não seria, de todo, evidente²⁴, uma vez que a Diretiva 2001/29 já excluía as reproduções transitórias, meramente instrumentais numa utilização legítima, aí incluindo a navegação em rede (*browsing*). Nesta perspetiva, a nova exceção teria por efeito limitar essa exclusão do ato de reprodução, permitindo aos titulares de direitos reservar a prospeção de textos e dados como utilização economicamente relevante. O mesmo vale, aliás, para os programas de computador, em que já antes era permitida a análise de programas objeto de ensino e de investigação, assim como a descompilação, sendo que a nova exceção de prospeção de textos e dados para fins de investigação científica não contempla agora a análise de programas de computador...

Por outro lado, ao abranger no exclusivo a prospeção de textos e dados, ainda que sujeita a reserva expressa, é discutível se a Diretiva 2019/790 não colocará um obstáculo significativo ao desenvolvimento da IA na Europa²⁵, em comparação com ambientes jurídicos menos rígidos,

²⁴ ELEONORA ROSATI, «An EU Text and Data Mining Exception for the Few: Would it Make Sense?», **Journal of Intellectual Property Law & Practice**, Vol. 13/6 (2018), p. 429-430.

²⁵ Thomas Margoni/Martin Kretschmer, «A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology», **GRUR Int.**, Vol. 71/8 (2022), p. 685-701 (“there should be no need for a TDM exception for the act of extracting informational value from protected works. The EU’s CDSM provisions paradoxically may favour the development of biased AI systems due to price and accessibility conditions for training data that offer the wrong incentives. To avoid licensing, it may be economically attractive for EU-based developers to train their algorithms on older, less accurate, biased data, or import AI models already

como o norte-americano, marcado por uma cláusula geral de *fair use* que confia aos tribunais a gestão casuística da adaptação da lei do *copyright* às necessidades do desenvolvimento tecnológico²⁶. Alega-se, inclusivamente, que a Diretiva teria sido mais uma “oportunidade perdida” para modernizar os direitos de autor face aos desafios da tecnologia, em especial da IA²⁷.

De todo o modo, o direito de autor da União Europeia é guiado pelo princípio do elevado nível de proteção da propriedade intelectual²⁸, reforçando o poder negocial dos titulares de direitos relativamente às empresas de «Big-Tech-Data» no que respeita a extração de valor por meio de prospeção de textos e dados das obras e outros materiais protegidos.

Na UE, os direitos de autor não são uma «terra de ninguém»: as obras protegidas por direitos de autor e outros materiais protegidos não podem ser utilizados livremente para alimentar novas tecnologias e as

trained abroad on unverifiable data”). Considerando a solução adotada “insensata” e “prejudicial à posição concorrencial da UE”, Tito Rendas, «Inteligência artificial, prospeção de dados e direito de autor», in **Propriedade Intelectual, Contratação e Sociedade da Informação. Estudos Jurídicos em Homenagem a Manuel Ohen Mendes**, coord. Dário Moura Vicente, Nuno Sousa e Silva, APDI/Almedina, 2022, p. 531-562, 559. Ver também Charlotte Gerrish, **European Copyright Law and the Text and Data Mining Exceptions and Limitations With a focus on the DSM Directive, is the EU Approach a Hindrance or Facilitator to Innovation in the Region?**, Uppsala University, 2019.

²⁶ VANESSA JIMÉNEZ SERRANÍA, «Datos, minería e innovación: ¿qvo vadis, Europa? Análisis sobre las nuevas excepciones para la minería de textos y datos», **Cuadernos de Derecho Transnacional**, Vol. 12/1 (2020), p. 247-258 (“es realmente necesario, a nuestro entender, replantearse seriamente a nivel europeo el impulso del llamado ‘flexible copyright’ en la aplicación de la regla de los tres pasos (threestep rules), especialmente en aras de permitir la actividad innovadora o transformativa” – p. 256). Sobre o tema, CHRISTOPHE GEIGER, «The Role of the Three-Step Test in the Adaptation of Copyright Law to the Information Society», **e-Copyright Bulletin**, 2007; Tito Rendas, **Exceptions in EU Copyright Law: In Search of a Balance Between Flexibility and Legal Certainty**, Kluwer Law International 2021; Eleonora Rosati, «Copyright reformed: the narrative of flexibility and its pitfalls in policy and legislative initiatives (2011–2021)», **Asia Pacific Law Review**, vol. 31/1 (2023), p. 33-54.

²⁷ Theodoros Chiou, «Copyright lessons on Machine Learning: what impact on algorithmic art?», **JIPITEC**, Vol. 10 (2019), p. 398-411, 411 (“the DSM Directive is a missed opportunity for true modernization of the European Copyright Law” – p. 411).

²⁸ Vd. por ex. Diretiva 2001/29, considerandos (4) e (9), Diretiva 2019/790, considerando (2).

empresas de IA. É claro que o desenvolvimento da IA é importante, mas não deve depender apenas dos direitos de autor, apesar das atuais proporções gigantescas das leis de direitos de autor.

A exceção vinculativa e obrigatória de prospeção de textos e dados para fins de investigação científica já abre caminho ao desenvolvimento da IA, incluindo parcerias com entidades privadas. A exceção geral vinculativa e supletiva de prospeção de textos e dados para outros fins onera os titulares de direitos autorais com a reserva de direitos, se não quiserem que as suas obras sejam livremente ‘mineradas’. Em vez de um direito de autorização, os direitos autorais são apenas um direito de oposição, pois o silêncio significa concordância com a autorização legal.

Provavelmente os desenvolvedores de IA não consideram este quadro suficientemente ‘liberal’. Mas, ao mesmo tempo, é possível que os titulares de direitos também se sintam parcialmente expropriados do seu direito de autorizar e por isso talvez uma compensação equitativa possa ser um bom remédio para eles no que diz respeito à prospeção para fins (e/ou por entidades) outros que não a investigação científica. A compensação equitativa foi excluída para a exceção obrigatória de prospeção, ficando em aberto se pode ser estabelecida relativamente à prospeção para outros fins²⁹.

Parece-nos que talvez possa ser um remédio equitativo para os detentores de direitos que não querem obstruir a mineração, mas que ao mesmo tempo gostariam de extrair algum sumo do fruto que criaram e que alimenta a IA gratuitamente, se não reservarem a atividade de mineração. A Lei da Rainha Ana para o incentivo à aprendizagem não significava que alguém aprenderia obtendo livros impressos gratuitamente...

A implementação portuguesa das exceções de prospeção não proporcionou uma compensação equitativa para a exceção geral e supletiva. Segue muito de perto a redação das disposições da diretiva, exceto no que

²⁹ Maria Victória Rocha, «Prospeção de texto e dados (Text And Data Mining) na Diretiva relativa aos Direitos de Autor e Direitos Conexos no Mercado Único Digital», *RRDDIS – Revista Rede de Direito Digital, Intelectual & Sociedade*, Vol. 2/4 (2022), p. 99-164 (destacando o “papel fundamental” destas exceções para “para impulsionar os *Big Data* e a inteligência artificial (IA) na União Europeia” – p. 160).

diz respeito à análise de software, que aparentemente não foi abordada pela legislação nacional, apesar do seu papel importante no desenvolvimento da IA e na aprendizagem automática (*machine learning*). Talvez o legislador nacional tenha concluído que a lei dos direitos de autor do software já previa regras equivalentes às da Diretiva no domínio do *software*.

REFERÊNCIAS

CHIOU, Theodoros. Copyright lessons on Machine Learning: what impact on algorithmic art?, *JIPITEC*, Vol. 10 (2019).

COMISSÃO EUROPEIA. **Livro Branco sobre a inteligência artificial – Uma abordagem europeia virada para a excelência e a confiança**, COM(2020) 65 final, Bruxelas, 19.2.2020.

DRAFT Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (16/5/2023), disponíveis em <https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf>. Acesso em: 15 set. 2023.

GEIGER, Christophe. The Role of the Three-Step Test in the Adaptation of Copyright Law to the Information Society, **e-Copyright Bulletin**, 2007.

GEIGER, Christophe; FROSIO, Giancarlo; BULAYENKO, Oleksandr. **The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects**, European Parliament, 2018.

GERRISH, Charlotte. **European Copyright Law and the Text and Data Mining Exceptions and Limitations With a focus on the DSM Directive, is the EU Approach a Hindrance or Facilitator to Innovation in the Region?**, Uppsala University, 2019.

GRIFFITHS, J.; SYNODINOU, T.; XALABARDER, R. **Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market**, European Copyright Society, 2022.

MARGONI, Thomas; KRETSCHMER, Martin. A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology, **GRUR Int.**, Vol. 71/8 (2022).

NORDEMANN, Jan Bernd; PUKAS, Jonathan. Copyright exceptions for AI training data—will there be an international level playing field? **JIPLP** 17/12 (2022).

PARLAMENTO EUROPEU. **Diretiva (UE) 2019/790 e do Conselho de 17 de abril de 2019** relativa aos direitos de autor e direitos conexos no mercado único digital e que altera as Diretivas 96/9/CE e 2001/29/CE.

PARLAMENTO EUROPEU. **Resolução de 20 de outubro de 2020**, sobre os direitos de propriedade intelectual para o desenvolvimento de tecnologias ligadas à inteligência artificial, para. 20 – Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_PT.html. Acesso em: 20 out. 2023.

PARLAMENTO EUROPEU. **Resolução de 3 de maio de 2022**, sobre a inteligência artificial na era digital. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_PT.html. Acesso em: 20 out. 2023.

PATTERSON, L. Ray; LINDBERG, Stanley W. **The Nature of Copyright: A Law of Users' Rights**, The University of Georgia Press Athens, 1991.

PEREIRA, Alexandre L. Dias. **Direitos de Autor e Liberdade de Informação**, Almedina, Coimbra, 2008.

RENDAS, Tito. "Are copyright-permitted uses 'exceptions', 'limitations' or 'user rights'? The special case of Article 17 CDSM Directive", **Journal of Intellectual Property Law & Practice** vol. 17 (2022).

RENDAS, Tito. **Exceptions in EU Copyright Law: In Search of a Balance Between Flexibility and Legal Certainty**, Kluwer Law International 2021.

RENDAS, Tito. Inteligência artificial, prospeção de dados e direito de autor. *In*: VICENTE, Dário Moura; SILVA, Nuno Sousa e. (Coords.) **Propriedade Intelectual, Contratação e Sociedade da Informação. Estudos Jurídicos em Homenagem a Manuel Ohen Mendes**, APDI/Almedina, 2022.

ROCHA, Maria Victória. Prospeção de texto e dados (Text And Data Mining) na Diretiva relativa aos Direitos de Autor e Direitos Conexos no Mercado Único Digital, **RRDDIS – Revista Rede de Direito Digital, Intelectual & Sociedade**, Vol. 2/4 (2022).

ROSATI, Eleonora. An EU Text and Data Mining Exception for the Few: Would it Make Sense?, **Journal of Intellectual Property Law & Practice**, Vol. 13/6 (2018).

ROSATI, Eleonora. Copyright reformed: the narrative of flexibility and its pitfalls in policy and legislative initiatives (2011–2021), **Asia Pacific Law Review**, vol. 31/1 (2023).

SENFLEBEN, Martin; BUIJTELAAR, Laurens. Robot Creativity: An Incentive-Based Neighbouring Rights Approach, **European Intellectual Property Review**, Vol. 42/12 (2020).

SERRANÍA, Vanessa Jiménez. Datos, minería e innovación: ¿qvo vadis, Europa? Análisis sobre las nuevas excepciones para la minería de textos y datos, **Cuadernos de Derecho Transnacional**, Vol. 12/1 (2020).

THE BRITISH COPYRIGHT STATUTE OF ANNE (1710). Disponível em: <https://www.copyrighthistory.com/anne.html>. Acesso em: 18 ago. 2023.

TRIAILLE, Jean-Paul; D'ARGENTEUIL, Jérôme de Meeûs; FRANQUENM, Amélie de. **Study on the legal framework of text and data mining (TDM)**, 2014.