

SET/OUT
2024



GEDIPE

NÚMERO
29

Associação para a Gestão Coletiva de Direitos de Autor
e de Produtores Cinematográficos e Audiovisuais



Inteligência artificial:
que futuro para o Direito do Autor nos EUA?

NEWSLETTER

EDITORIAL

ARTIGO DE FUNDO

EDITORIAL

POR PAULO SANTOS

DIRETOR-GERAL de GEDIPE

Como tem sido amplamente difundido (e comentado!) o Governo de Portugal tem um Plano de Ação para a Comunicação Social que se estrutura em quatro eixos principais: i) regulação, ii) serviço público concessionado, iii) incentivos e combate à desinformação e iv) literacia mediática.

Entre as diversas medidas anunciadas,² destacam-se a supressão gradual da publicidade na RTP, o apoio à distribuição de publicações periódicas em todos os concelhos do país e a oferta de 50% do valor das assinaturas digitais¹. Além disso, o Governo pretende dar incentivos ao jornalismo local e regional, financiar em 50% as assinaturas digitais de publicações de imprensa, apostar na formação para jornalistas e ainda incentivar a contratação de jornalistas.

No que diz respeito à projetada supressão da publicidade na RTP, ela será gradual, à razão de dois minutos por hora, em cada ano, até 2027.³ O que o Governo pretende com esta medida é que a RTP se concentre na sua missão de serviço público nas suas grelhas de programação retirando a dependência de receitas publicitárias, o mesmo será dizer, libertando-a de ter de assegurar níveis de audiência significativos. É suposto que o tempo dedicado à publicidade seja compensado com espaços de promoção de eventos e iniciativas culturais, à semelhança do que já sucede na RTP2 desde pelo menos 2003, data em que foi criada a CAV. A propósito, o Governo garantiu que não há planos para aumentar o valor dessa taxa, e que a perda de receita publicitária deverá ser compensada com ganhos de eficiência.⁴

A supressão de publicidade na RTP terá vários efeitos de médio e longo prazo no nível de investimento em produção audiovisual indepen-



dente, que suscitam preocupação ao setor: a eliminação da publicidade pode levar a uma redução na capacidade de financiar a produção de conteúdos diversificados, incluindo produções independentes.⁵ Por outro lado, a redução de recursos financeiros pode afetar a capacidade da RTP de oferecer uma cobertura nacional abrangente e de manter serviços essenciais, como a RTP Açores e a RTP Madeira, o que pode resultar numa menor oferta de programas e numa diminuição da produção de conteúdos nacionais,⁶ afetando direta e indiretamente trabalhadores diretos, freelancers e profissionais que trabalham em produções associadas à emissora. Isso inclui técnicos de cinema, atores, diretores de fotografia, entre outros podendo mesmo levar à perda de postos de trabalho em produtoras independentes, afetando a sustentabilidade do setor audiovisual como um todo,⁷ com um impacto profundo potencial na cultura e identidade portuguesa, bem como na sustentabilidade do setor audiovisual no país, que é um dos mais resilientes em termos comparativos com outros mercados audiovisuais, devido à preferência do público pela produção nacional.

Sucede, porém, que o resultado dessa medida pode não vir a ser o esperado pois a supressão de publicidade na RTP pode ter diferentes desdobramentos no mercado publicitário. Existem algumas possibilidades a considerar:

[1] <https://www.portugal.gov.pt/pt/gc24/comunicacao/noticia?i=governo-apresenta-plano-de-acao-para-a-comunicacao-social>

[2] <https://rr.sapo.pt/artigo/explicador-renascenca/2024/10/09/que-medidas-anunciou-o-governo-para-a-comunicacao-social/396917/>

[3] https://www.rtp.pt/noticias/economia/anunciado-plano-de-acao-para-a-comunicacao-social-governo-vai-retirar-publicidade-da-rtp_n1605766

[4] <https://rr.sapo.pt/noticia/politica/2024/10/08/governo-quer-fim-da-publicidade-na-rtp-sem-aumento-da-taxa-audiovisual/396669/>

[5] <https://sicnoticias.pt/pais/2024-10-14-fim-da-publicidade-na-rtp-ameaca-profissionais-do-audiovisual-alerta-associacao-de-imagem-664009ce>

[6] <https://www.jn.pt/6459060388/fim-da-publicidade-na-rtp-ameaca-profissionais-do-audiovisual-diz-associacao-de-imagem/>

[7] <https://observador.pt/2024/10/14/fim-da-publicidade-na-rtp-ameaca-profissionais-do-audiovisual-de-imagem/>



1. Redirecionamento para Televisão Privada

É possível que parte do investimento publicitário que deixará de ser canalizado para a RTP seja redirecionado para operadoras de televisão privada. Estas operadoras podem aproveitar a oportunidade para atrair anunciantes que buscam manter a visibilidade na televisão tradicional. Admite-se que possa ser este o objetivo principal da medida.

2. Aumento do Investimento em Publicidade Digital

Outra possibilidade é que os anunciantes optem por direcionar seus investimentos para plataformas digitais, como redes sociais e outros meios online. A publicidade digital tem-se mostrado uma alternativa eficaz para alcançar públicos segmentados e medir resultados de forma precisa. Se isso acontecer, quem ganhará serão as operadoras globais, a META, a Alphabet (Google), a dona do TikTok ou mesmo a X, de Elon Musk...

3. Diversificação dos Canais de Publicidade

Os anunciantes podem também optar por uma estratégia diversificada, distribuindo os seus investimentos entre diferentes canais, incluindo televisão privada, redes sociais, sites de notícias e outras plataformas digitais.

A decisão final dos anunciantes dependerá de vários fatores, incluindo o alcance e a eficácia dos diferentes canais, bem como as estratégias de marketing específicas de cada empresa. É pouco provável que cresça o budget para investir na televisão, a não ser que cresça nas formas comerciais mais inovadoras, por exemplo *product placement*.

Ou seja, a estratégia poderá ter efeitos adversos ao objetivo de dar sustentabilidade ao mercado, exceto se o Estado assumir, por outra via, o financiamento direto ao setor, que tem sido, aliás, significativamente preterido em comparação com o cinema de autor, tema que iremos abordar noutra ocasião.

Até lá, força para a criação audiovisual e votos de sucesso!

[8] <https://blog.swonkie.com/pt/publicidade-nas-redes-sociais-o-guia-completo-2024/>



POR VICTOR CASTRO ROSA

RESPONSÁVEL PELO GABINETE DE ESTUDOS
E RELAÇÕES EXTERNAS DA GEDIPE

Conforme anunciado no número antecedente desta Newsletter, fazemos hoje um ponto de situação sobre os mais importantes processos judiciais em curso nos Tribunais Norte-Americanos, cuja sorte poderá ditar os termos futuros do relacionamento entre o Direito de Autor e a Inteligência Artificial.

São 11, pelo menos, os casos que importa acompanhar com toda a atenção, sendo que os iremos apresentar por ordem cronológica da sua interposição, e após uma breve descrição do que está em causa, daremos nota dos aspetos que nos parecem mais importantes e que poderão ser decisivos, na nossa modesta opinião.

Hachette Book Group, Harper Collins, John Wiley & Sons e Penguin Random House contra Internet Archive

No dia 01.06.2020, os editores referenciados supra processaram a empresa Internet Archive (cujas iniciais são, curiosamente, IA) por alegada prática de digitalização (*scanning*) e empréstimo de cópias digitais de 127 livros sem a sua autorização.

A IA invocou a exceção de *Fair Use*, com base na sua missão de facultar o acesso universal ao Conhecimento, sem finalidades lucrativas. Além disso também foi invocada a "*First Sale Doctrine*" que postula que com a primeira venda ou comercialização autorizada deixam de ser invocáveis direitos de autor, e outros argumentos, como a caducidade e a ausência de danos provados. A decisão preliminar da primeira instância ocorreu a 24.03.2023, dando provimento ao pedido de julgamento sumário dos Autores e negando o pedido sumário de rejeição formulado pela Ré, não tendo considerado procedente a invocação de *Fair Use*.

O recurso, interposto pela Ré concluiu com a Decisão de 04.09.2024, da 2.ª Instância, na qual foi confirmada a decisão anterior e a responsabilidade direta por infração ao D.A.. Na decisão, foi determinante a não consideração do uso como "transformativo", porque não alterava o significado ou a expressão das obras sendo que o segundo fator qualificava as obras como criativas (gerando maior proteção) e o terceiro fator apontava para a reprodução integral das obras (permitindo um uso substitutivo das originais) e o quarto fator (impacto no mercado) também não favorecia a IA, mesmo limitado a

127 obras porque ao substituir as obras originais atuava como fator de desincentivo à criação intelectual.

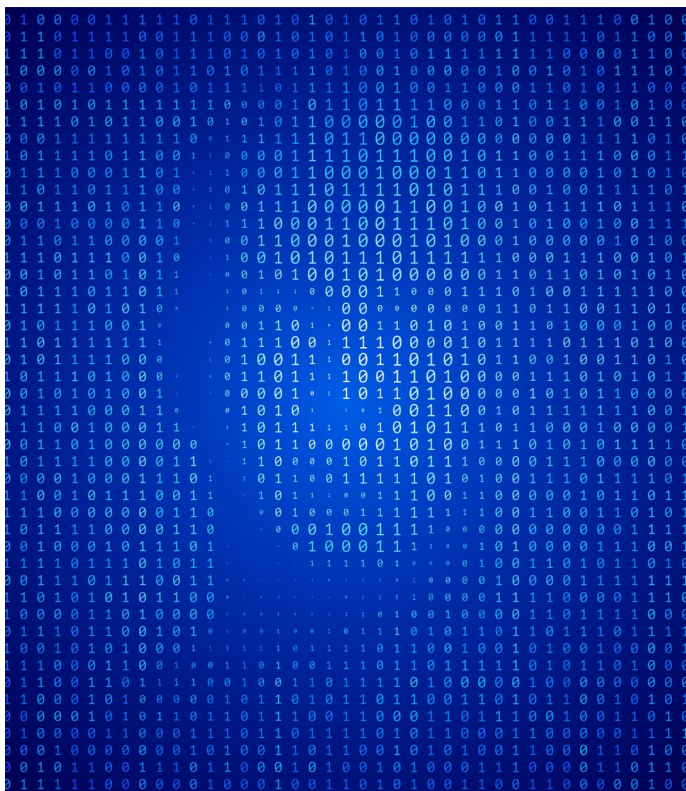
Esta característica diferenciava a situação face ao caso *Google Books* e *Hathi Trust*, que apenas disponibilizavam excertos ("*snippets*") não substitutivos. Os fins não lucrativos da IA não deixavam de ser comerciais devido ao pedido de donativos e às parcerias comerciais existentes que tornavam a utilização comercial.

Este caso constitui um precedente importante para os que iremos enunciar seguidamente, podendo desde já antecipar-se que, sempre que os Autores consigam demonstrar a prevalência ou a possibilidade do uso substitutivo que atente contra a exploração normal por parte dos titulares de direitos, não será considerado "*Fair Use*".



[1] Art.º 17 do Código de Direito de Autor os EUA, § 107, que abrange utilizações lícitas sem autorização tais como a crítica, o comentário, fins de informação, o ensino, a investigação científica ou académica e se baseia numa avaliação de quatro parâmetros, a saber: i) finalidade da utilização, nomeadamente o seu carácter transformativo; ii) a natureza das obras protegidas usadas; iii) quantidade e substância do uso; iv) efeito da utilização efetuada no mercado. A ponderação de cada fator no conjunto é feita caso a caso e varia bastante, uns podem pesar mais do que outros, e no final o resultado pode ser tangencial num sentido ou noutro.

John Doe (desconhecidos) contra GitHub, a Microsoft e a Open AI e outras



res não demonstraram uma identidade absoluta entre o código produzido e o original. Por aqui se vê qual é o critério que está a nortear os tribunais norte-americanos na apreciação deste tipo de pedidos.

Em 30.09.2024, aceitou o pedido de recurso interlocutório dos Autores, pelo que o caso segue agora para o *US Court of Appeals 9th Circuit*. Este caso poderá vir a ser importante para a definição das obrigações legais dos desenvolvedores de IA, com efeito nas ferramentas de codificação de IA e na indústria de software, determinando, por exemplo, a forma como o CoPilot utiliza código de Fonte Aberta disponibilizado por plataformas como a GitHub, todos pertencentes à Microsoft.

Na prática este processo serviu para os Tribunais dos EUA imporem às Rés uma série de obrigações que, na Europa, estão consagrados nos Regulamentos RGPD e de IA, entre outros, servindo de precedente para outros casos semelhantes, e inspirando, nomeadamente, o processo contra a Google que, como veremos, se baseia no uso de dados pessoais.

Esta ação foi instaurada em 03.11.2022 por um escritório de Advogados, como “*Class Action*” (na qual poderão tomar parte todos os que se encontrem na mesma situação dos Autores) com fundamento na infração dos respetivos direitos enquanto programadores, levada a cabo pelas ferramentas *Copilot* e *Codex*, envolvendo questões de privacidade, processamento de dados, violação de D.A. e pondo em causa as práticas da GitHub e os respetivos Termos & Condições contratuais. Foi requerida a decisão por júri.

Esta ação tem sobretudo a ver com a cópia de código de programação, protegido por D.A. enquanto “*software*”, mesmo no caso de Software de Código Aberto (sujeito a Atribuição e a Termos & Condições de Uso) mas também com a alegada falta de proteção dos dados pessoais dos utilizadores, para além de também haver acusações de Concorrência Desleal e violação dos próprios termos contratuais gerais subscritos pelas Rés e pelos utilizadores, nomeadamente no tocante à disponibilização dos respetivos dados pessoais a terceiros, relativamente às obrigações de atribuição de D.A. e por falta de transparência quanto aos termos e condições aplicáveis e também por enriquecimento sem causa.

Em junho de 2024 o juiz John S. Tigar rejeitou a maioria dos pedidos considerado que os Auto-



Sarah Anderson, Kelly McKernan, e Karla Ortiz contra Stability AI, DeviantArt Inc and Midjourney, Inc.



Este processo estabelecerá, certamente, um dos mais importantes precedentes no âmbito das imagens geradas por modelos IA que permitem produzir imagens e desenhos artísticos segundo o estilo ou com o traço característico de um/a artista gráfico/a. É mais uma “*Class Action*” instaurada pelas artistas gráficas acima identificadas, sendo que, neste momento, das iniciais, só uma continua na ação, Sarah Anderson, porquanto as outras não têm as suas obras registadas, o que, segundo as leis jusautorais norte-americanas, preclui o exercício judicial dos respetivos direitos. Assim, o processo está limitado a 16 obras desta artista, e outras de artistas que, entretanto, se juntaram à *Class Action*, estando em causa apurar se as ferramentas de conversão direta de texto (“*prompts*”) em imagens violam os D.A. das artistas ou se o resultado (“*output*”) é apenas um derivado dos respetivos trabalhos gráficos, sendo que, neste caso, também seria necessária a autorização.

Neste caso as Autoras solicitaram que o caso fosse julgado em tribunal de júri, o que permitirá uma decisão mais pautada pelo senso comum não tão exigente tecnicamente, o que pode ser uma forma estratégica de atingir um veredicto menos rigoroso do ponto de vista jurídico, por exemplo, deixando cair a exigência de total similitude entre as obras utilizadas para treinar o modelo de IA e os produtos que este gera.

A principal alegação consiste na utilização dos nomes das Autoras para as Rés promoverem os respetivos sistemas de IA generativa, publicitando a capacidade de produzir desenhos e artes gráficas “ao estilo” de cada uma das Autoras, sendo feita prova indiciária, pelo menos, de que as respetivas obras foram usadas no treino desses sistemas sem terem sido obtidos os respetivos consentimentos, afetando as suas reputações artísticas e permitindo a produção de imagens semelhantes às obras originais das Autoras. O Tribunal aceitou, como princípio de prova, os resultados de uma busca pelo nome da Autora no sítio “haveibeentrained.com” que permite saber se as suas obras foram usadas para treinar o modelo de IA generativa.

Os Artistas gráficos procuraram descrever, passo a passo, o processo de aprendizagem por parte dos algoritmos com base em métodos estatísticos e matemáticos simulando o raciocínio humano: os algoritmos são treinados a reconhecer milhares de milhões de imagens constantes de bases de dados obtidas na *Internet*, incluindo obras protegidas por Direitos de Autor e outros direitos de Propriedade Intelectual, sem que tenha sido obtida autorização por parte dos titulares. A dificuldade parece residir em que os modelos de IA generativa em causa apenas armazenam *versões comprimidas das imagens originais, não as pró-*

prias imagens em formatos perceptíveis para os humanos. No entanto, para as máquinas, é possível reconstituir imagens que, passando por um processo de “desintegração” de dados e depois de misturadas ou combinadas umas com as outras, nos seus elementos constitutivos essenciais, dão lugar ao produto final, por um processo de “limpeza de ruído”, em que apenas subsiste, no fim, uma imagem nova, fruto da “fusão” dos dados essenciais que integram cada uma das obras originais. A esta funcionalidade dá-se a designação de *Collage Tool* (colagem, amálgama).

No processo, os Artistas queixam-se também de serem eliminadas as informações para a gestão de direitos (metadados que lhes permitem identificar as respetivas criações e os usos interativos). A produção industrial de desenhos artísticos coloca sérios problemas de hipersaturação no mercado e dilui o reconhecimento e valor associados à criação original.

As dificuldades dos Artistas neste processo têm a ver com a forma como descrevem os atos violadores dos respetivos direitos, porque o processo de obtenção de imagens por IA não é assimilável a uma mera reprodução em suporte físico diferente, sendo que mesmo a recolha de obras originais na Internet pode não incluir nenhum ato de “cópia” deste tipo.

O juiz solicitou à Autora uma descrição mais pormenorizada do processo de treino para perceber se as Rés *criaram cópias das obras originais* ou se apenas usaram (e usam) *métodos estatísticos e matemáticos para capturar os respetivos conceitos* (a partir de imagens catalogadas/classificadas com metadados) o que pode não se configurar como infração ao D.A., *i.e.*, para haver infração, tem de haver reprodução da expressão original.

A própria alegação de responsabilidade objetiva (“*vicarious*”, ligada ao poder de impedir) carece, em rigor, de que seja verificada uma infração direta, não subsiste sem que alguém seja responsável por violação direta, pelo que é fundamental que haja esta última para se poder chegar à operadora que fornece a ferramenta de IA. Também no que se refere aos metadados (informação para gestão de direitos) o juiz solicitou a exemplificação concreta.

O juiz também encontrou deficiência na alegação de uso abusivo do nome dos artistas, porque não parece ter sido utilizado na publicitação ou promoção das ferramentas de IA nem achou que ficasse demonstrado que o uso desses nomes como “*prompt*” produzisse imagens tão similares às originais a ponto de confundir o público quanto à autoria daquelas. As alegações de concorrência desleal e de violação do contrato também foram consideradas improcedentes.

A defesa irá tentar demonstrar que nenhuma das Rés é responsável pela violação de D.A. uma vez que todas se limitaram a treinar os respetivos algoritmos com dados obtidos junto da empresa LAION-5B, que usou uma ferramenta criada em 2008 para “varrer” a Internet, mais de 3 biliões de *websites*, à procura de imagens e texto com imagens, não procedendo à “captura” ou “descarregamento” dessas imagens, mas recolhendo apenas dados da página, metadados e extratos de texto, armazenados como ficheiros WAT (*Web Archive Transformation*) que usam formato de texto *WebAssembly* para armazenar dados de páginas web de uma forma intermédia que permite o reagrupamento, mais tarde, em código binário de forma que um *web browser* ou outra ferramenta os possa ler e processar².

Uma vez que os ficheiros WAT não contém imagens em nenhum formato digital, apenas contém metadados que permitem à LAION ou outros utilizadores avaliar os metadados e dados de texto relativos a imagens disponíveis na Internet, sem nunca copiar ou descarregar essas imagens, pode ser mais difícil preencher o conceito de “reprodução”³.

Esta perspetiva é bastante controversa e já há quem se oponha a ela, considerando que a jurisprudência nos E.U.A., na União Europeia e no R.U. estende a noção de reprodução às *partes componentes de uma obra*⁴, e até mesmo aos fragmentos que podem ser reagregados para permitir uma utilização substancialmente idêntica a uma reprodução⁵. Por outro lado, há que considerar também a possibilidade de se tratar de uma forma de comunicação pública, considerando que a jurisprudência europeia tem vindo a qualificar como tal a efetivação de hiperligações para conteúdos quando existem proteções tecnológicas ao acesso⁶ ou quando esses conteúdos são reproduzidos noutros *websites*⁷.

[2] Num caso recente, o Tribunal Distrital de Hamburgo veio absolver a LAION de violação a D.A. aplicando ao caso a exceção de TDM contida no art.º 3.º da Diretiva MUD, não obstante se ter demonstrado que a recolha de dados serve para permitir a terceiros finalidades comerciais.

[3] Murray, M. (2023) Generative AI Art: Copyright Infringement and Fair Use, 26 SMU SCI. & TECH. L. REV. 259 disponível em https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4483539

[4] Acórdãos INFOPAQ, FAPL, ZIGGO e MIRCOM (TJUE)

[5] Rosati, E. (2024) Infringing AI: Liability for AI-generated outputs under international, EU, and UK copyright law, disponível em https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4946312

[6] Acórdãos SVENSSON, GSMEDIA e VG-BILD-KUNST (TJUE)

[7] Acórdão RENKHOFF (TJUE)

Richard Kadrey, Sarah Silverman, Michael Chabon contra META Platforms, Inc

Em 07.07.2023 deu entrada a “Class Action” movida por diversos autores de obras literárias contra a META Platforms Inc. devido ao modelo denominado LLaMa, que quer dizer *Large Language Model Meta AI* ou seja, a resposta da META ao Chat GPT da Open AI /Microsoft e ao BARD, agora chamado GEMINI, da Google. Em 12.09.2023, Michael Chabon instaurou outra ação semelhante que irá agora ser apensada e consolidada com a primeira.

As alegações dos Autores assentam na violação direta e indireta de D.A. pela utilização não autorizada das respetivas obras para treino do referido modelo, e a remoção indevida de Informação para a Gestão de Direitos. Adicionalmente, os Autores invocam a prática de atos de concorrência desleal, o enriquecimento sem causa e a atuação negligente ao lidar com material protegido por D.A. e ao omitir medidas apropriadas para proteger esses direitos. Por último, as alegações fazem eco de uma preocupação ética sobre as implicações da utilização de material protegido por D.A. no desenvolvimento futuro da IA.

O Tribunal, em 20.11.2023 deu parcialmente razão à defesa e deixou cair todas as alegações, com a possibilidade de aperfeiçoamento, menos na acusação de negligência, que rejeitou em definitivo. O Tribunal admitiu que o treino dos modelos de LLM pode constituir infração direta ao D.A., até porque a Ré não o negou na sua moção de rejeição. No entanto, considerou que para existir responsabilidade indireta na modalidade “*vicarious*”, os Autores teriam de alegar que o resultado produzido pelo modelo de IA continha a expressão protegida, na forma de reorganização, transformação ou adaptação das obras originais, citando a decisão do caso citado em 3) sustentou que são necessárias alegações de similitude substancial para apoio da teoria da infração por produção de obras derivadas sem autorização jusautorai. Ou seja, teria de ter sido alegado que os resultados produzidos pela aplicação “*incorporam, de alguma forma, uma porção*” dos respetivos livros. Não se pode contornar este requisito.

Em 11.12.2023 os Autores entregaram a petição melhorada na qual tiveram de exemplificar de forma extensiva as obras e os traços característicos do estilo de cada autor, bem como os temas típicos comuns às diversas obras de cada autor, ou aos diversos autores.

O Tribunal também rejeitou que o software LLaMA seja substancialmente similar à expressão concretamente adotada nas obras dos Autores,



e a mera extração de dados a partir de textos não é suficiente para configurar a infração na modalidade de obra derivada.

Também a alegação de supressão de dados para a gestão dos direitos fracassa porquanto os Autores não alegaram que a LLaMA tenha distribuído cópias dos livros, pelo que não faz sentido alegar distribuição sem metadados. Assim, o único aspeto que se mantém por explicar é se existe ou não reprodução das obras, até porque as leis federais de D.A. precludem a invocação de leis estaduais, como é o caso da concorrência desleal, do enriquecimento sem causa e de negligência.

Aguardam-se os próximos passos neste processo, sendo que a maior dificuldade que se antevê para os Autores, que têm falhado alguns prazos e pedido várias prorrogações ante o dilúvio de documentos apresentados pela META, irá ser a de estabelecer concretamente quais os atos que constituem infração direta ou indireta aos D.A., partindo do princípio de que o que está em causa é o estilo, a forma de escrever, a temática, etc., de cada Autor, *mas não a expressão concreta*, que é aquilo que é protegido pelo D.A, pelo que poderão perder, a não ser que o júri, que os Autores pediram para apreciar o pleito, lhes dê razão.

Jill Leovy, Jingna Zhang, Sarah Andersen et. al. v. Google LLC and Alphabet, Inc.

Esta “*Class Action*” foi instaurada inicialmente em 11.07.2023, seguindo-se outra queixa em 26.04.2024, sendo os processos apensados a pedido da Ré, atenta a identidade das questões.

A principal queixa nestes processos é a utilização indevida de dados pessoais dos Autores, ao abrigo da Lei de Proteção dos Consumidores da Califórnia, mas também é alegada concorrência desleal, negligência no manuseamento desses dados, invasão de privacidade, acesso indevido, receptação, privação de uso, enriquecimento sem causa, e violação dos direitos de autor, com remoção dos metadados para a gestão de direitos (DMCA), além de violação da Lei Antifraude Informática da Califórnia (CDAFA) e violação dos contratos estabelecidos com os responsáveis pelos websites de onde foi extratada a informação utilizada para treino do modelo Bard (atual Gemini) dos quais decorrem que os Autores são beneficiários.

Os Autores descrevem a forma como os seus dados pessoais (incluindo dados médicos, conversas no Gmail, informação financeira, e trabalhos criativos protegidos por D.A.) surgem nos serviços generativos da Google e invoca danos tais como a privação do controle dos respetivos dados pessoais, incluindo a interferência com a possibilidade de os mone-

tizar e a iminência de uso indevido para fins de desinformação e “*deepfakes*”. As práticas da Google são qualificadas como furto de larga escala, usando os dados pessoais para fins comerciais, distinguíveis da mera extração de dados (“*data scraping*”) sendo significativamente diferentes do que é aceitável mesmo para motores de busca, uma vez que os dados pessoais recolhidos foram inseridos (“*embedded*”) nos produtos IA sem o consentimento dos respetivos titulares, o que não tem qualquer apoio nos precedentes e configura um comportamento potencialmente criminoso para gerar lucros bilionários.

Os Autores responderam ao pedido de rejeição reiterando o detalhe das alegações. Apesar disso, não terão sido capazes de especificar que parcela de dados em particular foi utilizada ou como foram afetados pelo alegado uso indevido, o que levanta questões sobre a suficiência das alegações, sendo certo que já houve duas oportunidades para os Autores corrigirem as petições iniciais, por deficiências que deram origem a modificação dos pedidos, em lugar de um mero aperfeiçoamento e também devido à consolidação dos processos e concertação de estratégias processuais dos dois grupos de Autores dos dois processos apensados por decisão de 28.10.2024.



Concord Music Group, Inc. et al. contra Anthropic PBC

Este processo, instaurado a 16.11.2023 pode vir a ser o ponto de viragem na jurisprudência, desde logo pela circunstância de ser uma providência cautelar, contendo um pedido de imediata suspensão da violação continuada de D.A., mediante a proibição de utilização das respetivas obras para treinar o modelo de IA generativa Claude, datado de 16.01.2024 para além de proporem o licenciamento como forma de resolver a violação de D.A..

As autoras são todas editoras musicais, nomeadamente 8 grupos, os mais representativos da indústria musical, pelo que se trata de uma *Class Action* embora não formalmente.

O que caracteriza este processo é o volume considerável de evidências trazido aos autos, como prova de que os resultados do modelo Claude em matéria de letras de canções são substancialmente idênticos às obras cujos direitos exclusivos de reprodução e distribuição pertencem às Autoras, causando prejuízo significativo ao mercado, pondo em perigo o seu direito de controlo da distribuição e comprometendo as respetivas reputações.

São fornecidos diversos exemplos de respostas do Claude que são idênticas às obras protegidas por D.A., tornando impossível a ne-

gação dessa possibilidade, ou que só sucede caso seja especialmente procurada. Ou seja, pode acontecer mesmo que não solicitada, como quando se pede apenas uma poesia sobre um tema ou num determinado estilo. É também difícil negar que se trata de uma reprodução, ainda que efetuada a partir de uma representação abstrata, de memória, e não necessariamente a partir de uma cópia armazenada. A Ré até pode defender que é possível a coexistência do D.A. com a nova tecnologia e que respeita os direitos dos criadores, para evitar a ordem de suspensão, mas o certo é que não tomou a iniciativa de licenciar o uso que faz de obras musicais protegidas.

A posição da Anthropic é no sentido de não negar que tenha copiado massivamente obras protegidas sem a devida autorização, o que pode indiciar responsabilidade direta e indireta (contributiva) e não poderá proceder a negação de conhecimento de que tal uso é infrator. A alegação de "*Fair Use*" é fortemente prejudicada pela significativa parte reproduzida. Por outro lado, as alegadas proteções tecnológicas contra essas violações não funcionam.

Há um forte acentuar da vertente do licenciamento, em ordem a equilibrar a exploração económica efetuada pela Anthropic à custa da atividade das Autoras, Editoras Musicais. E enfatiza-se que tal não é acidental, mas um aspeto essencial da forma como a Ré opera.

Entretanto o processo, instaurado no Tennessee, onde fica a lendária cidade de Nashville, irá ser transferido para a Califórnia do Norte por acordo das partes, na sequência de moção da defesa, uma vez que a sede da Ré é naquele outro Estado e é aí que é coordenado o treino do Claude, aí se encontrando também a maioria das testemunhas. É expectável que a deslocação do processo não seja desfavorável à posição das Autoras, que representam o grosso da Indústria Musical e a decisão pode servir de precedente para a necessidade de obter licenciamento junto dos titulares de D.A. para o treino de modelos de IA generativos.

Note-se que este caso configura um claro exemplo de como o resultado gerado pelo modelo pode ser, efetivamente, uma reprodução fiel da obra original, capturando, portanto, a expressão protegida pelo Direito de Autor. Dificilmente se admite que fracasse, até porque também foi pedido que a avaliação do mérito do caso fosse feita por um júri.



Michael Chabon and David Henry Hwang contra Open AI e Microsoft

Em 04.12.2023, deu entrada mais uma *Class Action*, desta feita instaurada por vários Autores literários contra a Open AI e a Microsoft, por violação direta do DMCA através das cópias efetuadas sem autorização das obras literárias dos Autores, para fins de treino dos modelos de IA generativa das Rés, nomeadamente, o Chat GPT e o Bing AI, agora "CoPilot".

Os Autores qualificam as produções geradas pela AI como *obras derivadas das originais*, demonstrando as similitudes ou identidade textual, que não pode ser explicada de outra forma senão mediante alguma forma de reprodução não autorizada a partir da *Internet*.

Pedem uma indemnização legalmente estabelecida, mas também a compensação dos danos que venham a ser provados e, em último caso, invocam enriquecimento sem causa.

O destino deste processo será, com toda a probabilidade, o mesmo que tiverem os outros casos em que a causa de pedir seja a violação de direitos de autor pela utilização de obras literárias para treino dos modelos de IA, com as dificuldades inerentes à demonstração da plausibilidade de haver "similitude substancial" entre as obras originais e o resultado da IA.

New York Times contra Microsoft e Open AI



Este processo foi interposto a 27.12.2023, após falhar uma abordagem negociada do NYT à Open AI e à Microsoft, sendo que estas não consideraram oportuno licenciar os conteúdos.

O jornal NYT alega que a Open AI utiliza o conteúdo das respetivas edições para treinar a IA, sem autorização, violando o D.A., extraindo o conteúdo de milhões de edições, eliminando a Informação para a Gestão dos Direitos (metadados) para facilitar a infração continuada, e disseminando resultados do processo generativo que incluem cópias e obras derivadas através de plataformas como Bing Chat (atual CoPilot) em benefício do seu produto de IA e em concorrência direta com os produtos do NYT, prejudicando a exploração económica. A infração é direta, mas implica responsabilidade indireta "contributiva" e "vicarious". O NYT enfatiza a importância de sustentar uma atividade

que é essencial à democracia e ao combate à desinformação, que prolifera na Internet e nos vários meios de comunicação. Não foi reclamada ainda uma indemnização mas poderá sê-lo se proceder.

Neste processo, as probabilidades de sucesso do NYT prendem-se com a capacidade de demonstrar que é possível gerar exatamente os mesmos textos publicados pelo jornal, se a instrução dada for precisa, o que torna a IA uma ferramenta substitutiva da compra ou subscrição dos próprios jornais e liquida o 1.º, o 3.º e o 4.º argumentos de Fair Use, ou seja, o uso não é transformativo, a quantidade usada é substancial e afeta o NYT do ponto de vista comercial ou de mercado. Do lado da Microsoft e da Open AI tudo depende do peso do argumento de que não pode haver direitos exclusivos e monopólio de factos e de ideias.

Daily News contra Microsoft e Open AI

Neste processo, instaurado a 30.04.2024, as alegações são substancialmente semelhantes ao Processo instaurado pelo NYT, mas com algumas especificidades e abrangendo diferentes tecnologias:

Os Autores referem concretamente exemplos de artigos jornalísticos indevidamente apropriados, focando-se no conteúdo produzido por jornais locais e nos respetivos direitos exclusivos de exploração que são postos em causa não apenas pela criação e distribuição dos resultados obtidos por virtude do treino com cópias não autorizadas mas também pela *memorização de porções significativas dessas obras*, o que permite a “produção” ou “geração” de *cópias praticamente idênticas quando solicitado pelos utilizadores*. Ou seja, a reprodução a partir da “memória artificial” gerará e disponibilizará ao público “obras derivadas”, mas sem informação para a gestão de direitos, *i.e.* sem os metadados que as identificam, removidos intencionalmente para impedir a deteção de novas infrações.

Outra alegação consiste no dano para a imagem e diluição das marcas dos jornais, na medida em que as informações divergentes e ilógicas (“*Hallucinations*”) podem gerar desinformação e até mesmo difamação de pessoas e entidades concretas, o que levanta problemas éticos e sugere uma nova área de litigiosidade relativa à responsabilidade das empresas tecnológicas na gestão dos respetivos produtos. Os autores qualificam essa responsabilidade como direta, mas também contributiva e objetiva, devido ao envolvimento significativo na comercialização de modelos que violam os direitos de autor.

Neste processo, durante o seu depoimento, o fundador da Open AI Sam Altman terá assumido que a empresa necessita de conteúdo protegido por direitos de autor e conexos para poder fornecer um produto comercialmente viável, o que se pode relacionar com outras afirmações conhecidas do mesmo responsável (embora não neste processo), segundo as quais estaria disposto a pagar as compensações que viessem a ser exigidas.

Os Autores propuseram que este processo fosse apensado à ação interposta pelo NYT, o que as Rés também requereram em 13.06.2024 e 14.06.2024, para evitar decisões contraditórias em processos que assentam nos mesmos factos e na mesma legislação.

É esperável que o desenlace deste processo facilite a via do licenciamento e do “*enforcement*” do Direito de Autor, mas também que seja clarificado o âmbito do “*Fair Use*” em ambiente digital, assegurando o respeito e a compensação dos titulares de direitos.

Outra consequência possível será o desenvolvimento de modelos colaborativos entre a indústria tecnológica e os editores de publicações noticiosas com vista à criação de produtos de valor acrescentado que beneficiem ambas as partes e também se espera conquistar o apoio do público para a causa do jornalismo de qualidade e mudar a atitude dos consumidores em ordem a preferirem apoiar diretamente os editores do que usar os conteúdos gerados pela IA que não compensam adequadamente os criadores originais.



Andrea Bartz, Charles Graeber and Kirk Wallace Johnson contra Anthropic PBC

Este processo foi instaurado a 08.08.2024 e envolve jornalistas e autores literários que se queixam de violação dos respetivos direitos de autor exclusivos (reprodução e distribuição) mediante a utilização das suas obras para treino dos modelos de IA da Ré Anthropic, com prejuízo para a exploração normal pelos Autores, alegado como irreparável por forma a justificar uma medida cautelar imediata e permanente, para além da reclamação de uma indemnização legalmente fixada e encargos decorrentes do processo. A acção é também configurada como *Class Action*, pelo que outros Autores nas mesmas condições poderão associar-se à mesma. Por outro lado, esta condição implica um cuidado acrescido com a justiça das decisões e os valores acordados, bem como uma rigorosa avaliação das propostas e justificações detalhadas para as decisões a tomar, incluindo por parte do júri.

Uma das alegações essenciais neste caso é de que a Ré terá efetuado o “*download*” de versões pirata das obras dos Autores, a partir de *websites* ilegais e efetuou reprodução dos mesmos, utilizando-os como material de treino, para permitir respostas “humanizadas”.

A defesa alega que os Autores não demonstraram adequadamente que as respetivas obras estavam incluídas no conjunto de dados utilizado para treino dos modelos de IA, e ataca a natureza heterogénea dos autores envolvidos, para efeito da sua qualificação como classe. A complexidade desta questão processual, que constitui um incidente autónomo, será alvo de muita atenção, considerado o precedente que pode gerar para futuras ações deste tipo.

Dow Jones & Company, Inc. e NYP Holdings, Inc. contra Perplexity, Inc.

Este processo foi instaurado a 21.10.2024 pela empresa proprietária dos jornais “The Wall Street Journal” e “New York Post” e, além da usual alegação de violação de D.A., assenta muito particularmente na falsa atribuição de notícias e afirmações, com evidente desprestígio para as marcas dos jornais e atentando contra a respetiva credibilidade, que é o seu principal ativo, o que compromete a exploração comercial e, conseqüentemente, faz perigar a continuação do trabalho jornalístico, que perde o seu incentivo financeiro. No passado, os titulares de direitos não foram afetados por motores de busca que apenas incentivavam a descoberta dos conteúdos, já a IA funciona como produto substitutivo.

Uma das principais alegações é a de que a Ré oferece um produto substitutivo, desviando os subscritores do produto original, e espalhando a desinformação com notícias falsas que faz passar como sendo originalmente atribuíveis aos jornais, invocando-os abusivamente.

Neste processo também é particularmente acentuada a falta de vontade da parte da Ré em obter uma licença para a utilização que faz dos conteúdos publicados pelo seu modelo IA, ao contrário de outras empresas de IA (é referida a recente parceria da Open AI com a News-corp, que permite a utilização de conteúdos jornalísticos, quer como resumo quer como ci-

tações) sendo referida a perda de oportunidade de parceria comercial de sucesso que permitiria à Ré explorar outros tipos de produtos, havendo todo um mercado de oportunidades comerciais a explorar em parceria, com um enorme potencial de benefícios comuns, sem que sejam as empresas de IA a ditar os respetivos termos, que é o caminho apontado pela Autora, ao invés de uma apropriação abusiva dos conteúdos que confunde os próprios consumidores e prejudica o mercado, minando a confiança na informação.



Conclusões



Como refere Caterina Sganga citada no recente Estudo do Observatório Europeu do Audiovisual “*AI and the Audiovisual Sector: Navigating The Current legal Landscape*”, pag. 28, nota 99 (trad. livre): “*ao contrário das máquinas, os humanos não recordam os verdadeiros objetos mas apenas as ideias conceptuais desses objetos*”. Esta diferença é que torna mais difícil que alguém seja capaz de reproduzir um quadro ou uma partitura musical apenas de memória, sem ter perante si uma cópia física ou um registo de imagem. Já os modelos de IA generativa conseguem reconstituir um texto ou uma imagem porque sabem exatamente onde encontrá-lo, ou seja, guardaram na sua supermemória as coordenadas que permitem torná-lo presente, *sem precisarem de o reproduzir novamente*. No entanto, são capazes de o *reconstituir*, se para tal solicitados ou mesmo fortuitamente. Se os tribunais se ativerem a um critério estrito de comparação e procura da “*similitude substancial*” para determinarem a existência de violação ao Direito de Autor, poucos serão os casos que terão decisões positivas, ainda que os júris possam ser mais compreensivos.

Se for adotado um conceito mais abrangente do âmbito dos direitos exclusivos, quer seja pelo alargamento do conceito de reprodução à reconstituição integral ou quase-integral a partir de uma supermemória, quer pela de-

terminação judicial de que a referenciação à localização na *Internet* é também uma forma de comunicação ao público (em particular quando a disponibilização *online* não é livre ou obedece a regras contratuais restritivas) ou pela tutela dos direitos de personalidade ou sobre os dados pessoais, como sucede em alguns dos pleitos referenciados, então talvez haja alguma esperança para os casos em que não seja possível encontrar e demonstrar uma identidade absoluta entre o “*input*” e o “*output*” como resultado do processo criativo operado pelos modelos de IA generativa.

Supomos que os magistrados profissionais nos EUA terão alguma dificuldade em adotar decisões com alguma imaginação face a uma nova realidade que desafia os limites da tutela pelo Direito de Autor, sendo certo que também estão a rejeitar outras formas de tutela como é o caso da concorrência desleal, por razões de “preclusão” por leis federais.

Já na UE, onde é admitida a reserva expressa contra a extração de texto e dados (TDM) pode ser que baste aos titulares de direitos exercerem formalmente um “*opt-out*” (não obstante ser uma inversão do Princípio Geral de Direito de Autor, que não depende de formalidades) para ser considerada uma situação de infração qualquer ato de treino de modelos de IA.



GEDIPE

**Propriedade:**

GEDIPE - Associação para a Gestão Coletiva de Direitos de Autor e de Produtores Cinematográficos e Audiovisuais

Pessoa coletiva n.º 504 229 290

Sede Social:

Av. Infante D. Henrique, n.º 306,
Lote 6, 1.º andar
1950-421 LISBOA

Diretor:

António Paulo Santos
paulo.santos@gedipe.org

Redator Principal:

Victor Castro Rosa
victor.rosa@gedipe.org

Edição, imagens e composição gráfica:

ADBDB Comunicare - Consultores
Associados, Lda.
Av. da Igreja 42, 1700-035 Lisboa