

LEGAL ALERT

ADC IDENTIFICA UM CONJUNTO DE QUESTÕES CONCORRENCIAIS NO ACESSO A DADOS NA IA GENERATIVA

Em setembro de 2024, a Autoridade da Concorrência (AdC) adotou um *Short Paper* intitulado “Concorrência, IA Generativa e Dados”. O documento surge na sequência do seu *Issues Paper* de 2023, dedicando-se, agora e em particular, a um conjunto de questões jusconcorrenciais suscitadas pelo acesso e importância dos dados no setor da Inteligência Artificial (IA) generativa.

A centralidade da IA generativa, capaz de gerar novos conteúdos – como texto, imagem, sons ou vídeos –, não oferece dúvidas e vem suscitando desafios em vários domínios do Direito. No Direito da Concorrência, os riscos situam-se em diferentes momentos da cadeia (treino e inferência), relativamente aos seus vários participantes (*developers* dos modelos, fornecedores de componentes, serviços de computação), e respeitando aos vários inputs essenciais que, além dos dados, incluem poder computacional, conhecimentos técnicos avançados e financiamento. Os dados têm impacto não só no treino dos sistemas, mas também na sua verificação e monitorização, pelo que são relevantes ao longo de toda a operação de sistemas de IA.

O *Short Paper* da AdC foca-se em apenas um dos inputs estratégicos para o desenvolvimento da IA generativa – os dados –, e aborda:

1. A evolução entretanto ocorrida no setor da IA generativa (de dados publicamente disponíveis para dados proprietários);
2. Os riscos associados a acordos de exclusividade e ao acesso preferencial a dados;
3. O papel limitado dos dados sintéticos na garantia de contestabilidade e, bem assim,

4. As vantagens dos modelos de IA de código aberto na mitigação dos efeitos de escala associados ao pré-processamento de dados.

I. Questões jusconcorrenciais suscitadas pelo licenciamento de dados

A propósito do acesso a dados, a AdC sinaliza a íntima conexão entre o crescimento da IA generativa e a utilização de dados públicos (por exemplo, dados disponíveis em repositórios de acesso livre na Internet como a *Wikipedia*), sendo estes, pelo menos numa primeira fase, cruciais para o treino de modelos de IA.

Na sequência, porém, de incertezas quanto ao enquadramento aplicável em matéria de propriedade intelectual (PI), agravadas por uma sucessão de reações e de litígios opondo criadores de conteúdos e titulares de direitos aos fornecedores de IA generativa (de que o caso *New York Times c. OpenAI* é exemplo), houve lugar a uma profusão da celebração de acordos de licenciamento de dados, para fins, ora de treino, ora de *grounding* (exigindo uma utilização recorrente dos dados).

Este movimento translativo de dados publicamente disponíveis para dados proprietários suscita, segundo a AdC, riscos jusconcorrenciais associados (i) à criação ou reforço de barreiras à entrada e à expansão no mercado e, bem assim, (ii) ao reforço do poder de mercado de empresas incumbentes. Em ambos os casos, agravados pelo emprego de cláusulas de exclusividade e, bem assim, por práticas de acesso discriminatório (preferencial) a dados.

Perante os riscos de práticas abusivas por empresas com posição dominante no mercado (proibidas pela Lei da Concorrência nacional e pelas disposições dos Tratados da União Europeia), a AdC avança a necessidade de agilizar os processos de licenciamento de dados, através:

1. Da disponibilização de dados através de API abertas;
2. Do agrupamento de licenças em pacotes; e
3. Da adoção de estruturas de preços de pagamento pay-as-you-go, para assim evitar os efeitos de escala.

II. A insuficiência dos dados sintéticos para assegurar contestabilidade

A AdC reconhece, no seu *Paper*, que os dados sintéticos, isto é, artificialmente gerados e passíveis de utilização posterior no treino de novos modelos de IA generativa, poderão mitigar as barreiras à entrada e os custos de aquisição de dados, contribuindo para a contestabilidade do mercado. Salienta, também, as vantagens destes dados em matéria de privacidade, proteção de informação confidencial e garantia de diversidade.

Em contraste, porém, as insuficiências dos dados sintéticos (em termos de desempenho, fiabilidade, erros de geração e enviesamentos) não permitem afastar os riscos suscitados pelas vantagens competitivas dos fornecedores de IA de finalidade geral, e decorrentes da possibilidade de acesso a dados reais.

III. A importância do pré-processamento de dados no desenvolvimento da IA generativa

O último ponto assinalado no *Paper* refere-se ao pré-processamento de dados (*data filtering* ou *data selection*), qualificado pela AdC como um «passo essencial no treino de qualquer modelo de IA e um fator diferenciador chave». Tendo em conta a diversidade de técnicas de pré-processamento de dados, a AdC sinaliza a necessidade de escolha de um *mix* ótimo, que inclua a remoção de dados de baixa qualidade e duplicados, e a mistura de dados de fontes diferentes.

Porque também aqui se verifica a necessidade de acesso a inputs chave, como recursos computacionais, tempo e pessoal especializado, a que acresce a própria experiência associada àquele conjunto de técnicas, a importância dos modelos de IA em código aberto e respetiva documentação-transparência são particularmente assinaladas pela AdC como vias para mitigar os efeitos de escala e os consequentes riscos de concentração do mercado.

IV. Conclusão

O *Short Paper* da AdC inscreve-se no âmbito da sua missão de *advocacy*, não sendo nem vinculativo nem indicativo de um qualquer curso de ação no domínio da IA generativa. Desde logo, e como salientado pela AdC, a evolução do mercado da IA generativa está, ainda, dependente da

resolução de algumas questões em matéria de proteção de dados, propriedade intelectual e, até, *governance* da própria IA.

O âmbito do *Paper* é, também ele, limitado, não se debruçando sobre outros riscos jusconcorrenciais, associados, por exemplo, a outras infraestruturas essenciais (e respetivos prestadores); ao papel das parcerias e investimentos minoritários, envolvendo *developers* de modelos e empresas ativas, por exemplo, no setor da *cloud*; aos mercados laborais; e à integração vertical.

Mesmo no que se refere ao acesso a dados, algumas questões justificariam um tratamento mais aturado, porventura numa iniciativa conjunta com outras autoridades com competência setorial (por exemplo, o acesso a dados dos utilizadores; o cruzamento entre as regras da concorrência, da concorrência desleal e os direitos de propriedade intelectual, etc.), ou através da densificação do potencial abusivo de algumas das práticas assinaladas como suspeitas.

Ainda assim, no estágio de desenvolvimento atual, é compreensível que as questões sejam mais do que as respostas. Com este *Paper*, a AdC demonstra estar atenta e vigilante dos riscos de um setor em crescimento, sinalizando, através de uma abordagem próxima, as abordagens que devem ser ensaiadas para mitigar as questões jusconcorrenciais relativas ao acesso a dados.

Trata-se, repare-se, de uma iniciativa não isolada, devendo, além de outros relatórios, estudos de mercado e análises por várias autoridades nacionais da concorrência, ser particularmente assinalado o *Competition Policy Brief* da Comissão Europeia, intitulado “Competition in Generative AI and Virtual Worlds”.

A [equipa de Europeu e Concorrência da Morais Leitão](#), em conjunto com o [ML Digital Cluster – Artificial Intelligence](#), continua a monitorizar de perto os desenvolvimentos horizontais e setoriais passíveis de impactar diretamente nos seus clientes.

Gonçalo Rosas
Inês F. Neves
Tiago Ramos Cunha

Esta publicação é meramente informativa, não constituindo fonte de aconselhamento jurídico nem contendo uma análise exaustiva de todos os aspetos dos regimes a que se refere. A informação nela contida reporta-se à data da sua divulgação, devendo os leitores procurar aconselhamento jurídico antes de a aplicar em questões ou operações específicas. É vedada a reprodução, divulgação ou distribuição, parcial ou integral, do conteúdo desta publicação sem consentimento prévio. Para mais informações, contacte-nos por favor através do endereço comunicacao@mlgts.pt.