

## LEGAL ALERT

### ADC IDENTIFIES A SERIES OF COMPETITIVE ISSUES IN DATA ACCESS IN GENERAL AI

In September 2024, the Competition Authority (AdC) adopted a [Short Paper](#) entitled ‘Competition, Generative AI and Data’. The document follows on from its 2023 [Issues Paper](#), now focusing in particular on a set of fair competition issues raised by data access and its significance in the generative Artificial Intelligence (AI) sector.

The centrality of generative AI that is capable of generating new content – such as text, image, sounds or videos – is beyond doubt and has been raising challenges in several areas of the Law. In Competition Law, risks are situated at different points in the chain (training and inference), in relation to its various participants (model developers, component suppliers and computing services), and in relation to the various essential inputs which, in addition to data, include computing power, advanced technical knowledge and financing. The data has an impact not only on the training of the systems, but also on their verification and monitoring, and are therefore relevant throughout the entire operation of AI systems.

The AdC Short Paper focuses on only one of the strategic inputs for developing generative AI – the data – and addresses:

1. Developments in the meantime in the generative AI sector (from publicly available data to proprietary data);
2. The risks associated with exclusivity agreements and preferential access to data;
3. The limited role of synthetic data in ensuring stability, as well as
4. The advantages of open source AI models in mitigating the scale effects associated with data preprocessing.

## **I. Fair competition issues raised by data licensing**

With regard to data access, the AdC highlights the close connection between the growth of generative AI and the use of public data (for example, data available in free access repositories on the Internet such as *Wikipedia*), which are, at least initially, crucial for the training of AI models.

However, following uncertainties regarding the applicable framework for intellectual property (IP), exacerbated by a series of reactions and disputes between content creators, rights-holders and providers of generic AI (such as the case of the *New York Times* and *Open AI*), there has been a proliferation of data licensing agreements, whether for training or grounding (requiring recurrent use of data).

According to the AdC, this movement of publicly available data to proprietary data raises fair competition risks associated (i) with the creation or reinforcement of barriers to entry and expansion in the market, as well as (ii) with the strengthening of the market power of incumbent companies. In both cases, these have been exacerbated by the use of exclusivity clauses and by the practice of discriminatory (preferential) access to data.

In view of the risks of abusive practices by companies with a dominant position in the market (prohibited by the Portuguese Competition Law and by the provisions of the Treaties of the European Union), the AdC advocates the need to streamline data licensing processes, through:

1. Making data available through open APIs;
2. Grouping of packaged licensing; and
3. Adopting pay-as-you-go price structures to avoid the effects of scale.

## **II. The insufficiency of synthetic data to ensure stability**

The AdC acknowledges in its Paper that synthetic data, i.e. artificially generated and subject to later use in the training of new generative AI models, may mitigate entry barriers and data acquisition costs, thereby contributing to market stability. It also highlights the advantages of these data in terms of privacy, protection of confidential information and ensuring diversity.

In contrast, however, the insufficiencies of synthetic data (in terms of performance, reliability, generation errors and biases) do not allow the risks arising from the competitive advantages of general purpose AI providers, resulting from the possibility of access to actual data, to be removed.

### **III. The importance of data preprocessing in the development of generative AI**

The last point highlighted in the Paper refers to data pre-processing (data filtering or data selection), viewed by the AdC as an ‘essential step in the training of any AI model and a key differentiating factor’. In view of the diversity of data preprocessing techniques, the AdC points out the need to choose an optimal mix, which includes the removal of poor quality and duplicate data, and the mixing of data from different sources.

Because here too there is a need for access to key inputs, such as computing resources, time and specialised personnel, in addition to the experience associated with that set of techniques, the importance of open-source AI models and their documentation-transparency are particularly highlighted by the AdC as means of mitigating the effects of scale and the consequent risks of market concentration.

### **IV. Conclusion**

The AdC's Short Paper is part of its advocacy mission, and is neither binding nor indicative of any course of action in the field of general AI. From the outset, and as pointed out by the AdC, the evolution of the generative AI market is also dependent on resolving some issues around data protection, intellectual property and even the governance of the AI itself.

The scope of the Paper is also limited and does not focus on other fair competitive risks associated, for example, with other essential infrastructure (and respective providers); the role of minority partnerships and investments involving developers of models and companies active in, for example, the labour markets; and vertical integration.

Even with regard to access to data, some issues would justify closer processing, perhaps through a joint initiative with other authorities with sectoral competence (e.g. access to user data; cross-

referencing between competition, unfair competition and intellectual property rights, etc.), or by increasing the potential for abuse of some practices flagged as suspicious.

Even so, at this stage of development, it is understandable that there will be more questions than answers. With this Paper, the AdC demonstrates that it is alert and vigilant as regards the risks of a growing sector, highlighting, through a close analysis, the approaches that should be tested to mitigate fair competition issues relating to data access.

This is, it should be noted, a non-isolated initiative and, in addition to other reports, market studies and analyses by various national competition authorities, the European Commission's Competition Policy Brief entitled 'Competition in Generative AI and Virtual Worlds' is of particular significance.

The [European and Competition team at Morais Leitão](#), together with the [ML Digital Cluster - Artificial Intelligence](#), continues to closely monitor horizontal and sectoral developments that may directly impact its customers.

Gonçalo Rosas  
Inês F. Neves  
Tiago Ramos Cunha

This publication is purely for information purposes and is not meant to be a source of legal advice, nor does it contain a comprehensive review of all aspects of the law and legal frameworks it refers to. The information contained herein refers to the date of first publication. Readers must seek legal advice before applying it to specific issues or transactions. The contents of this publication may not be copied, disclosed or distributed, fully or in part, without prior consent. For more information please contact us at [comunicacao@mlgts.pt](mailto:comunicacao@mlgts.pt).