Generative AI and Competition Policy

2024 Florence Digitalisation Summer Conference

17 June 2024

Key-note Speech

Nuno Cunha Rodrigues

Thank you for the invitation and for the opportunity to share the views of the Portuguese Competition Authority on the emerging field of AI.

It is for me an honor to to be the keynote speaker at this important event and to be back to the EUI, not only as President of the AdC, but also as a Professor of Law.

This is an exciting time to be a competition enforcer.

Since late 2022, with the launch of ChatGPT, digital markets are undergoing a major upheaval. Generative AI is a clear example of disruptive innovation and it is rapidly advancing.

It is able to produce new content in diverse forms (text, images, video, audio) with a wide range of possible usages.

Also, Generative AI may increase the world GDP by 7% in the next decade[1] and can transform markets, business models and the global economy.

The potential of AI is already palpable.

Nevertheless, **we are still in the beginning of the road.**

There is a lot of experimentation happening and the sector is still addressing hallucinations by language models.[2]

---

[1] Goldman Sachs (2023), "Generative AI could raise global GDP by 7%", https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html.

[2] They often do not distinguish fact from falsehood and provide nonsensical, and sometimes dangerous information. The rollout of AI Overview by Google is one of many examples of this. It has been documented citing satirical news as fact or giving dangerous advice to users – such as claiming poisonous mushrooms are safe. https://www.iflscience.com/googles-ai-overviews-is-returning-bizarre-

Nowadays, the key competition concern focuses on the potential bottlenecks in AI development and implementation.

In fact, if these bottlenecks are controlled and exploited by a small number of firms, competition concerns may arise.

And this is a real risk.

The development and implementation of AI models require (i) data; (ii) computing power and (ii) know-how.

Competition authorities have been closely following the developments in the sector, via policy documents, as well as through inquiries into partnerships and investments, which have been prolific across the AI value chain.

The Portuguese Competition Authority is attentive to this new sector.

The AdC has published, in November last year, an Issues Paper on generative AI and competition.

Since then, we have continued monitoring the developments of the sector.

As competition authorities we must fully realise the benefits and contribute to the potential of AI, ensuring the proper incentives for firms to innovate.

Having said this, I will introduce the roadmap of my intervention.

First, I will go (i) over **each of the key AI inputs** highlighting a few trends in the sector and, then, (ii) pinpoint some of the competition concerns.

At the end I will touch upon on the (iii) regulatory developments at a global scale that may impact the AI sector.

Let me start with the key inputs of AI.

---

and-even-dangerous-results-74375https://www.washingtonpost.com/technology/2024/03/18/ai-mushroom-id-accuracy/. More examples here: https://gizmodo.com/google-search-ai-overview-giant-hallucination-1851499031 and https://www.theverge.com/2024/5/23/24162896/google-ai-overview-hallucinations-glue-in-pizza

Access to data and data partnerships

First, data and access to data.

AI models live and die by the data they are trained on.

Data can be any form of digital content as generative AI works by replicating the respective patterns.

Because data is a key input for generative AI, we shall answer three key questions: First, we need to understand how to best protect data that is subject to **Intellectual Property (IP) and privacy rights.**

Second, **how to make sure developers have access to large, diverse and high-quality datasets**.

Third, the extent to which training datasets are **substitutable** or **non-replicable**.

In the past, these three questions might not have been as relevant.

AI models were mostly trained on datasets that were publicly available, such as large internet archives of webpages, which could be accessible by everyone.

Nowadays, **a significant portion of this training data is subject to IP rights**.

As business models and commercial applications of AI begin to mature, the holders of IP rights start demanding compensation for the use of their data.

Whether AI developers may or may not use this data without authorization from IP holders is an open question, and will, of course, depend on the legislation in each jurisdiction.

Still, **this is ultimately a legal risk for AI developers** that I do know will be addressed in this conference

It is worth noting that, currently, it is difficult to know exactly what training data AI models are using.

The technical reports no longer provide the details they used to[3].

[3] E.g., OpenAI gave many details about its training data for GPT-3, but not GPT-3.5 or GPT-4. Likewise, Meta was very open about the training data it used in Llama 1 and Llama 2, but for Llama 3 it gives no

While developers may be giving less details in order to protect trade secrets from their competitors[4], they are probably also trying to avoid this legal risk.

In fact, there have been several lawsuits regarding the unauthorized use of copyrighted material[5].

As such, answering to our first question, I should point out that **AI developers have begun entering into data licensing agreements with IP holders.**

These agreements include:

- publishers, like Associated Press[6] or Alex Springer[7];
- stock image repositories, like Shutterstock[8]; and
- social networks, such as Reddit[9] or Stack Overflow[10].

These are some of the partnerships we know about.

Still, it is likely there are many out of public view.

This is effectively **creating new markets for data** having into account that these data sources are akin to producers of data, leading to the importance of our second question, which was how to make sure developers have access to large, diverse and

---

information, claiming solely that the data is from publicly available sources. Other models, such as Phi-3, from Microsoft, the models from Mistral AI or Gemini from Google also give little to no information about training data.

[4] The CEO of Mistral says so in this forum: https://huggingface.co/mistralai/Mistral-7B-v0.1/discussions/8

[5] For example, the New York Times[5] and the Author's guild sued OpenAI. Getty Images – a stock images repository – and a group of artists sued Stability AI. This list is by no means exhaustive. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html. Informative discussion on the case here: https://hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/. https://authorsguild.org/news/ag-and-authors-file-class-action-suit-against-openai/. https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/.

[6] https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a

[7] Owner of publications such as Bild, Politico or Business Insider. https://openai.com/index/axel-springer-partnership/

[8] https://investor.shutterstock.com/news-releases/news-release-details/shutterstock-expands-partnership-openai-signs-new-six-year

[9] https://blog.google/inside-google/company-announcements/expanded-reddit-partnership/ https://www.wired.com/story/reddit-ipo-filing-s-1-cofounder-alexis-ohanian-huffman/ https://www.wired.com/story/reddits-sale-user-data-ai-training-draws-ftc-investigation/

[10] Stack overflow is a platform where computer programmers may ask and answer questions. It is widely used. https://stackoverflow.co/company/press/archive/openai-partnership

high-quality datasets.

A first concern relates to who gets to decide on data access. Presently the decision tends to be left to the incumbents.

The most interesting case is perhaps that of social networks.

Because users want to share content and interact with other users, social networks are subject to strong network effects, thereby creating a tendency towards concentration.

A clear example are the recent news regarding the intention of Facebook and Instagram to leverage users' information to train its AI systems, taking us back to privacy rights concerns.

This may also give them substantial market power in data licensing. **As so, data sources should have incentives to license their data to a large array of AI developers.**

Still, conceptually, **there may be risks granting exclusive access** if monopoly rents are shared with these data sources, potentially creating first mover advantages.

Moreover, breaches to IP, or even privacy, rights can lead to competition harms. These may be used to attain, exploit or entrench market power, or raise entry barriers.

Also, in what concerns our last question, regarding the markets for data there is the issue on the substitutability of data.

It is relevant to establish the type and the amount of data necessary for the performance of the model.

To better understand this dimension, it may also be necessary to consider the impact of combining different data from distinct sources, as well as whether the remarkable ability of the models to generalise can accommodate limited access to data.

A last point on data I would like to mention is **monitoring data.**

Developers track their own models by observing users and use this information to

train and optimize future versions of the models.

These may also generate network effects, as developers with the largest user bases are observing a richer set of behaviours and have the best means to conduct A/B experiments.

Access to cloud computing and specialised hardware

The second key input for generative AI is computing power.

In fact, training and deploying generative AI models requires significant computing power.

As in the case of data, this is especially true for foundation models.

AI developers may access computing power either via their own infrastructure of specialised hardware, such as computer chips specialised for AI, namely GPUs and "AI accelerators" (or NPUs), or they can contract cloud computing services.

Given the upfront costs of building own infrastructure, the only option for most AI developers seems to be to rely on cloud computing. Those who have that infrastructure, such as Google or Microsoft, have had a significant competitive advantage prior to the "AI race", as they are key players in the cloud computing business.[11]

Meta is perhaps an exception, with an investment in its own infrastructure aimed at rivalling the capability of some cloud providers[12].

This investment has been estimated to be valued at $30 billion in 2024 alone. Nonetheless, Meta investment also entails a plan to acquire in 350 000 high-grade GPUs from Nvidia[13].

In this context, many AI developers have been seeking to secure partnerships with

---

[11] Microsoft and OpenAI have announced a $100 billion investment in a supercomputer for AI. https://archive.is/OvM8o

[12] https://engineering.fb.com/2024/03/12/data-center-engineering/building-metas-genai-infrastructure/

[13] This figure has been estimated and reported in an interview to Yann LeCun, the chief AI scientist at Me, which has confirmed the report. https://www.youtube.com/watch?v=6RUR6an5hOY. The $30 billion figure should already include a discount, as the H100 GPUs from Nvidia have been reported to cost up to $40 000 each. See here and here.

cloud computing providers, to secure access to compute.

Microsoft, Amazon and Google have formed cloud computing partnerships with many AI developers, including OpenAI, Mistral, Anthropic, Cohere or Stability AI[14].

These partnerships may also entail:

- investments by cloud providers in the AI developer;
- bespoke collaboration between teams;
- licensing the use of AI models by the cloud provider; or
- distribution agreements of AI models.

However, both the markets for cloud computing – with Microsoft, Amazon and Google as the major players – and computer chips specialised for AI, namely GPUs– with Nvidia standing out – are very concentrated.

As a result, **most AI developers may be extremely dependent on their upstream suppliers.**

It also highlights **how incumbents in cloud and computer chips may be in a unique position to enter and establish a foothold in AI markets**.

Cloud providers may have the ability and incentives to privilege compute towards one AI model, namely their own.

Another factor that might also play a role in the incumbents' advantage is the energy demand of these infrastructures.

Indeed, they are energy intensive creating a further dilemma in terms of the imperatives of sustainability.

The energy required to run AI tasks is accelerating, representing an annual growth rate between 26% and 36%.[15]

---

[14] The most notable partnership of this sort is the one between Microsoft and OpenAI. And this example actually highlights the need and value of computing power. It has been reported that a significant portion of Microsoft's multi-billion-dollar investment in OpenAI has been paid off in computation time in Microsoft's cloud. https://archive.is/Yx6EC https://www.semafor.com/article/11/18/2023/openai-has-received-just-a-fraction-of-microsofts-10-billion-investment

[15] https://www.weforum.org/agenda/2024/04/how-to-manage-ais-energy-demand-today-tomorrow-and-in-the-future/

This can be a further argument towards the use of shared cloud computing resources.

In fact, also in this area there are partnerships being established, namely by the cloud providers: Let me remind you of the agreement celebrated by Microsoft with Helion Energy[16].

Access to know-how and experimentation

The third major input in generative AI is know-how.

Developers require specialised personnel to develop, implement, experiment and train AI models.

So far, this is has proved to be a scarce resource.

So much, that there seems to be a scramble for specialised personnel in the AI sector, with developers trying to attract talent from other developers.

As a result, firms may have incentives to restrict the mobility of their employees, for example via non-compete clauses or restrictive off-boarding agreements[17].

One competition concern also relates with the potential for anticompetitive agreements, such as no-poach, which may introduce misallocations in labour markets.

Given the key relevance of AI experts, these types of agreements may also harm competition in downstream product markets, by limiting knowledge spillovers and harming innovation.[18]

---

[16] https://world-nuclear-news.org/Articles/Google,-Microsoft-and-Nucor-team-up-on-clean-energ

[17] There have been reports that the former employees at OpenAI, for example, are subject to restrictive off-boarding agreements. These contain both NDAs, to protect confidentiality and trade secrets, but they also impede former employees from criticizing OpenAI. If breached, former employees could lose the equity they hold in OpenAI. https://www.vox.com/future-perfect/2024/5/17/24158478/openai-departures-sam-altman-employees-chatgpt-release https://www.vox.com/future-perfect/351132/openai-vested-equity-nda-sam-altman-documents-employees

[18] https://www.concorrencia.pt/sites/default/files/Issues%20Paper%20Labor%20Market%20Agreements%20and%20Competition%20Policy%20-%20final.pdf

Indeed, there is evidence that labour mobility is positively related to productivity, namely in sectors intensive on R&D, such as AI.[19]

It is thus important that competition authorities remain vigilant and promote best practices in AI labour markets.

Experimentation is also key.

This is related to know-how but includes, specifically, the organizational knowledge firms build over time.

The development of AI models is often described by practitioners as an "art".

Building a model requires many small decisions that cumulatively may have a large impact on performance.

It requires both time from already scarce employees, but also collecting a lot of data that may never be used and expending computing power and energy in each of the experiments.

For example, Meta mentions how it performed "extensive experiments" to evaluate the best ways to mix data from different sources when building its training dataset for the Llama 3 model.[20]

Experimentation and know-how are thus key for the optimization of AI models, requiring large capital investment.

Major competition concerns

The characteristics related to each input identified – data, computing power, expertise and experimentation – may generate **strong scale effects**, **network effects and a lack of multi-homing**.

As a result, **digital incumbents** and **first movers may accumulate competitive advantages** that can lead to entrenched market positions and potentially harmful competitive behaviour.

---

[19] See, e.g., https://academic.oup.com/chicago-scholarship-online/book/28640/chapter-abstract/239174673?redirectedFrom=fulltext

[20] https://ai.meta.com/blog/meta-llama-3/

All in all, there may be a **tendency towards concentration** in the AI sector and in adjacent markets.

This tendency can arise, for instance, from the **exclusivities in access to data**, which can lead, on the one hand, to the foreclosing of access to data or, on the other hand, rendering it exceedingly difficult to obtain a license from all the right holders.

Another concern may emerge as a consequence of the **incentives to vertical integration**, **agreements and partnerships**, which can undisputably also lead to efficiencies.

Nonetheless, the integration of AI in other incumbents' products, such as cloud computing, the dependency of AI developers and the nature of the partnerships may raise competition questions, namely the risk of self-preferencing or whether some the agreements may qualify as transactions subject to merger control.

Self-preferencing can be particularly worrying regarding foundation models owners that enjoy market power, given the ability and incentive to give their own Generative AI models privileged or preferential access downstream.

Access to AI models and open-source

Indeed, access to AI models, and in particular foundation models, plays a central role in the dynamics of competition in downstream markets for Generative AI.

A more generalized deployment of the existing foundation AI models will likely increase the number of applications for AI, benefiting AI in a quicker and more extensive manner.

However, both having access and ensuring the quality of that access are key concerns.

A key part of the quality of access depends on how open the foundation model is. This openness is ultimately a matter of degree.

At one extreme, only the developer has access to the foundation model.

At the other extreme, the model is effectively open source, namely by being subject to a license compliant with the Open-Source Definition (OSD) as set by the Open Source Initiative[21].

This includes, for example, making model files available[22] and allowing the use or distribution of fine-tuned models.

Most of the cases fall in-between the two extremes.

Without direct access to the model files, the number of ways the developer can do fine-tuning or further adapt the model in any way is greatly diminished.

As a result, the downstream AI developer faces more limitations in materializing an application for AI, thus potentially harming innovation.

Still, **there may be incentives for foundation model developers make model files available** – a key feature of openness.

Open-source fosters a network of developers around the foundation model, boosting innovation and ideas on how to improve the model.

The open-source model may also become a standard in the sector, such that the costs of developing on top of the model are lower relative to the competition.

And releasing models in open-source creates a talent pipeline for the developer itself, as people learn and gain experience on these models.

There are, however, a number of risks or limitations to this openness.

First, the open-source format does not preclude anticompetitive practices, as they may still be used to strengthen the market power of its developer in an adjacent market.

---

[21] This is a standards organization, which has coined the term "open-source" and its main activity has been to set the standards for what is to be deemed "open-source", https://opensource.org/about https://opensource.org/osd

[22] "Model files" refers to the model weights. These are files which contain the architecture, hyperparameters and the trained parameters of the model. Using these files, one can do fine-tuning, provided one has the expertise and computing power to do it, independently of the foundation model developer. To do fine-tuning effectively one typically needs additional documentation, to know how it was built and how it works.

Second, it may be difficult to monetize open-source foundation models and these may not necessarily stay open.

Third, developers may introduce restrictions on the possible uses and applications of the model at any time.

For instance, the Llama models by Meta have been widely considered open source.

Nonetheless, their licenses place limitations on the developers who build AI applications on top of Llama.

It establishes that if applications built using Llama models reach more than 700 million active users, the AI developer needs to obtain a license from Meta, which may or may not be granted, at the sole discretion of Meta.

Moreover, AI developers cannot use Llama to develop other language models (except for Llama itself or models based on it).

There are also restrictions in some use cases due to safety concerns[23].

In fact, because of these restrictions, the Open-Source Initiative, who sets the standards for what is "open source", considers the Llama models not to be a "true" open source[24].

Despite these limitations, open foundation models are likely to introduce competitive discipline on the most performant models.

Distribution and integration of AI models

For open-source foundation models, or any other model, to effectively compete; **end users and software developers must find it easy to access.**

This is a question of distribution or integration of AI models into existing digital services.

Foundation models, for example, are often distributed via cloud platforms, such as

[23] , namely related to pharmaceutical products, or in operating critical infrastructure, transportation technologies or heavy machinery.

[24] The article by OSI refers only to Llama 2, but what is written is also applicable to Llama 3, given the similarity of the licenses. https://opensource.org/blog/metas-llama-2-license-is-not-open-source

Microsoft Azure.

In fact, many of the partnerships between cloud providers and AI developers are also about ensuring one's foundation model is available in the cloud service for all its customers.

If such models are not available on major cloud platforms, downstream developers will find it difficult to use them.

This is especially true for proprietary models.

So far, it seems the major cloud providers are open to integrating many models into their service.

But risks remain.

Incumbents with large digital ecosystems may have the ability and the incentive to accelerate the development of their own AI by integrating it with their own products, at the expense of competing AI.

On the contrary, such tying strategies could go the reverse way, where a dominant position in AI is leveraged to protect existing digital products in the digital ecosystem.

This risk is especially relevant given the generality of foundation models and the possibility of having one or very few foundation models for everything.

Intersection with the DMA

Beyond competition policy, other regulations may have an impact on competition in the AI sector.

The DMA is likely on top of the list.

It is a regulation born out the lessons learned by competition authorities about digital markets.

It is not about any digital market.

It is focused on multi-sided markets, namely in markets where there is a platform between end users and business users.

Nonetheless, it has been argued[25] that the **DMA may apply or impact on generative AI services.**

On the one hand, generative AI models may be considered a core platform service in the future, falling within Article 2(2) or, specifically, under Article 2(12) of the DMA, in accordance with the definition of virtual assistants[26].

On the other hand, the DMA may apply to generative AI because of its integration into existing core platform services [Article 2(2) of the DMA].

Search engines complemented with AI are the most notable example.

For instance, by displaying answers generated by language models to users' queries alongside search results.

This is already available in Bing and Google.

Moreover, many obligations in the DMA, namely the ones set out in Article 5, refer other services provided by the gatekeeper.

Because some of the major players in AI are also gatekeepers, these obligations may apply to AI automatically.

A clear example is the obligation provided in the number 2 of Article 5 of the DMA, that gatekeepers obtain consent to combine personal data from a core platform service to any other service.

This said, the extent to which AI services are within the scope of the DMA is still largely open.

It will depend, namely, on how AI services are defined – for example, as either integral to a core platform service or as a separate service.

In fact, the DMA may not be sufficient to fully cover the risks associated with the development and deployment of Generative AI.

---

[25] See, among others, the European Parliament "Report on artificial intelligence in a digital age", adopted on 22 March 2022, 2020/2266(INI), in particular paragraphs 66 and 168.

[26] Article 2(12) of the DMA: "*'virtual assistant' means a software that can process demands, tasks or questions, including those based on audio, visual, written input, gestures or motions, and that, based on those demands, tasks or questions, provides access to other services or controls connected physical devices.*"

AI regulation and global reach of AI

Generative AI specifically raises a wide range of concerns beyond competition policy, including, for example, issues related to privacy, IP rights, as previously mentioned, safety and fundamental rights.

Regarding this last concern – fundamental rights -, AI may represent a risk to the democratic process and the public discourse, including by creating fake news or personalized disinformation.

In India[27], for example, AI has been heavily used by political candidates, and deep fakes abound, albeit many with positive ends (e.g. candidates speaking in local language or dialect).

To harness AI's unique opportunities and mitigate its risks, various strategies for the sector are being developed worldwide.

The EU AI Act is one of those strategies[28].

It classifies different AI systems according to their risk, and lays down obligations that AI systems must abide, including some specific requirements for AI deemed to be high risk[29].

The AI act also envisages **specific obligations towards foundation model developers and downstream providers of AI applications.**

In the US, **President Joe Biden issued an executive order** on 'Safe, Secure, and Trustworthy Development and Use of AI', issued in October 2023[30].

The UK has adopted a principle-based approach, having established a framework for existing regulators to interpret and apply within their sector-specific domains.

---

[27] https://www.wired.com/story/indian-elections-ai-deepfakes/

[28] https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[29] Most AI applications currently in use have minimal risk and are not subject to restrictions. At the other end of the scale, some applications are deemed to pose unacceptable risks, and therefore banned. These include the untargeted scraping of CCTV to build facial recognition databases.

[30] It builds on earlier work related to combating algorithmic discrimination and the securing of voluntary commitments from major US companies to drive safe, secure, and trustworthy AI development.

In China, new rules on Gen AI entered into force in August 2023, the so-called "Interim Measures on the management of generative AI services" and impose obligations on services offered to the public.

In fact, also in 2023 the G7 issued a voluntary AI code of conducts, setting international guiding principles, applicable to all AI actors across the value chain, to promote safety and trustworthiness in advanced AI systems.

As such, there are different approaches on how and when to regulate AI.

The EU opted for a regulatory and human-centric approach, providing transparency obligations which aim to incentivise the development of reliable content and the application of security controls.

Perhaps it is still too early to access the exact impact of the different approaches across distinct jurisdictions, as many implementation acts or legislations may follow.

Even though these measures are not about competition, they may nonetheless have an impact in the future landscape of AI markets not only by introducing compliance costs, hitting the hardest smaller firms, who may not afford to both comply with complex rules and invest in new technologies,

But also due to the risks associated with regulation not keeping pace with the technological progress and business models.

Indeed, it is an equilibrium in which intervention and regulation needs to find as not to unnecessarily stifling competition and innovation.

One interesting debate relates to the mentioned open-source models.

Some argue in favour of open models because they foster innovation and enhance transparency within the AI sector.

Others argue that open-source models present considerable security risks, such as facilitating disinformation and should be approached with cautious.

Prioritizing safety at this extent over competition or innovation will very likely create barriers to entry and expansion, and potentially restrict the availability of open-source models.

Indeed, according to Article 53(2) or Article 54(5) of the EU AI Act open-source foundation models are exempted of certain obligations, unless they are monetizing it or put into service as high-risk AI systems or an AI system that falls.

Similar debates relate to training data.

A pro-competition regulatory approach regarding training data would highlight the importance of access to data for competition. As a result, all non-personal data would be potentially licensable.

Overall, international cooperation, for instance within the OECD, and a better understanding, including the close monitoring of the industry development, are key to tackle not only risks associated with safety, privacy and fundamental rights, but also to preserve the process of dynamic competition and innovation.

After all, I remind you that AI is very likely to transform the global economy.

Conclusion

It is time to conclude.

Some call AI as a "black box."[31] It is our job to understand how it works and mitigate the long-term risks.

Generative AI has the potential to create new markets and disrupt existing ones. In fact, it also seems to be a poster child for all the challenges digital markets pose to competition.

As a competition enforcer, we want firms to experiment new ideas, business models and applications. Our goal is to overcome externalities, remove barriers, so that firms have the incentives and means to innovate.

Therefore, we must remain vigilant to ensure that competition is right from the get-go, ensuring that markets are open and contestable, to the benefit of consumers and the economy.

[31] https://time.com/6980210/anthropic-interpretability-ai-safety-research/