bitkom

Competition in GenAI

Bitkom's Feedback to the European Commission's Call for Contributions on competition in generative AI

What are the main components (i.e., inputs) necessary to build, train, deploy and distribute generative AI systems? Please explain the importance of these components

The main components necessary to build, train and distribute generative AI systems (or Foundation Models (**»FMs«**)) are data, compute capacity, and technical expertise. Data is needed at both stages of training: (i) pre-training, where data is used to build the FM's knowledge; and (ii) fine-tuning, where the FM's accuracy is improved through dedicated training. Further, the Data has to be of high quality (no systematical bias, no harmful speech, no old/overcome knowledge, no personal data, no political bias e.g.).

Significant compute capacity is required to train FMs. Because of the size of the model and the amount of data used to train the model, FMs require a significant number of AI accelerator chips (such as graphic processing units – **»GPUs«**) installed in large data centers (public/private cloud or on-premise). Some FM developers have their own AI accelerator chips called tensor processing unit (**»TPU«**).

FM developers can also turn to cloud service providers (public/private cloud) to have access to this type and scale of compute capacity. These cloud providers need access to enough GPUs or other AI accelerators to be able to service their FM customers. Some FM developers also use on-premise hosted GPUs or other AI accelerators. Further Alternatives such as publicly owned supercomputers are also relevant/emerging (e.g. ↗AI Innovation Pack, EuroHPC and related activities).

FM development also requires a combination of talent and technical expertise. This includes data scientists and engineers, machine learning skills, programming, mathematics and statistics. The skills and expertise needed may vary depending on the type and complexity of the FM (e.g. expertise or knowledge of image data, natural language processing, or text/speech/video/music generation).

What are the main barriers to entry and expansion for the provision, distribution or integration of generative AI systems and/or components, including AI models? Please indicate to which components they relate.

Regarding Provision: The development/provision and distribution of generative AI is a highly dynamic space. Its significant growth and the number and diversity of players, including many start-ups, evidence the dynamics of the market in general.

However, a barrier to entry are high costs and a lack of financial support on the provision layer. The capital market does not usually invest in long-term projects. Private and public investments can lower barriers to entry and expansion of generative AI developers.

In addition, the availability of FMs through APIs and open-source licenses enables new entrants to enter and scale quickly. Yet it can be challenging for new entrants to catch-up with existing providers of FM without having to rely on these APIs and open-source licenses.

Regarding distribution: At the distribution level, various deployment models are relevant (on-premise, private cloud, public cloud, etc.). In general, the main barriers are therefore the same as in this market in general in the context of the deployment of other applications, although there are of course specifics in the context of generative AI. Both training and pure inference are relevant here in terms of distribution, and there are different hardware requirements (GPU, CPU) in each case.

Regarding integration: It exists a broad spectrum and therefore intense competition among general and specialized providers and service providers for IT system integration. With the rise of generative AI, another field of activity has been added for these providers.

What are the main drivers of competition (i.e., the elements that make a company a successful player) for the provision, distribution or integration of generative AI systems and/or components, including AI models?

Regarding provision: In general, it should be noted that the discussion about the extent to which the performance of a generative AI model should be measured in a meaningful way is ongoing. in the context of the negotiations on the AI Act, this was also a central issue. The agreement[1] has strengths and weaknesses. This issue must be analyzed and discussed further in the context of further market development.

It is difficult to predict what kind of competition issue may emerge in this market horizontally. Not only because it is fast-evolving but also because it is still at an relatively early stage. It is unclear at this stage how forthcoming regulation will affect competition and to what extent potential concerns will be addressed by forthcoming regulation or regulation already in force, in particular by the DMA.

It should also be noted that it is foreseeable that there will be a central differentiation and thus competition on the market via sector and domain-specific fine-tuned models. A purely horizontal view is therefore not sufficient.

Regarding distribution: At the distribution level, various deployment models are relevant (on-premise, private cloud, public cloud). In general, the main barriers are therefore the same as in this market in general in the context of the deployment of other applications, although there are of course specifics in the context of generative AI. The upcoming implementation of the Data Act will also facilitate the switching of customers from one cloud provider to another and enhance multi-cloud solutions that will foster competition. Both training and inference are relevant here in terms of distribution, and there are different hardware requirements (GPU, CPU) in each case.

Regarding integration: See answer in question 2 in this context.

[1] specifically Article 52a (2) "cumulative amount of compute used for its training [...]) in combination with Annex IXc.

Which competition issues will likely emerge for the provision, distribution or integration of generative AI systems and/or components, including AI models? Please indicate to which components they relate.

In general, the answers from question 3 are also relevant with regards to question 4 (future competition which will likely emerge).

In addition, it is difficult to predict what kind of competition issue may emerge in this market. Not only because it is fast-evolving but also because it is still at a relatively early stage. It is unclear at this stage how forthcoming regulation will affect competition and to what extent potential concerns will be addressed by forthcoming regulation or regulation already in force, in particular by the existing competition rules (Art. 101 and Art. 102 of TFUE) and the DMA. It is important to give existing and forthcoming regulation time to unfold its full impact in an effective manner, and also not to worsen the existing competition structure.

How will generative AI systems and/or components, including AI models likely be monetised, and which components will likely capture most of this monetization?

There is potential for monetization across the various layers of the AI technology stack:

AI accelerator chips/GPUs: As they are the baseline for training/development of Generative AI systems in general there is a high monetization potential in the current and future market.

Compute resources. Cloud service providers (public/private) sell cloud capacity (i.e. access to GPUs installed in their data centers) to FM developers. Offering on-premise GPU infrastructures is also a business model.

The provision of compute resources is also a business model in the context of a) use-case/domain-specific fine tuning and b) deployment of generative AI systems (GPUs/CPUs in the cloud or on-prem).

Data. Datasets can be proprietary and may be licensed to FM developers for the purposes of training a model. Other data sets are open source and are publicly available for download.

The business models of providing generative AI Models are also central. Both as-a-service provision models and licensing models play a role here. Further differences and specifics exist with regards to the question of whether models are proprietary or open source (question 6).

Do open-source generative AI systems and/or components, including AI models compete effectively with proprietary AI generative systems and/or components? Please elaborate on your answer.

Regarding performance: Ad hoc, we see no systematic difference between open source and proprietary generative AI systems and/or components when it comes to the question of whether they compete effectively against each other. There are high-performance models in both segments.

Regarding business models: Open source and proprietary generative AI systems and/or components models follow different business model logics. It remains to be seen how the business models in both areas will develop and what this means for the market shares/relevance of proprietary and open source models.

What is the role of data and what are its relevant characteristics for the provision of generative AI systems and/or components, including AI models?

See answer to questions 1-3 in this context.

Developers have access to a significant amount of data on the open internet and in publicly available datasets. In addition, the EU data strategy, including the creation of common European data spaces, will make even more data available for access and reuse. Success will not always be linked to the volume of data but will include higher quality data, better algorithms or the use of less data in a smarter way to be more cost-effective.

What is the role of interoperability in the provision of generative AI systems and/or components, including AI models? Is the lack of interoperability between components a risk to effective competition?

Generative AI systems and components are generally interoperable on the horizontal layer. Chatbots and virtual assistants are often created by combining different FMs from different developers.

With regards to vertical integration, the role/question of interoperability is more complex and should be monitored as the market develops (compare question 9). The Data Act will contribute to interoperability among cloud services' providers for the benefit of customers.

Do the vertically integrated companies, which provide several components along the value chain of generative AI systems (including user facing applications and plug-ins), enjoy an advantage compared to other companies? Please elaborate on your answer.

The role & implications of vertical integration in the Generative AI value chain should be closely monitored as the market develops.

Many technology companies active in the generative AI field operate at different levels of the AI stack. However, most of them still need to collaborate or enter partnerships with other parties to be able to successfully bring their products or services to the market. For example, some companies provide computing infrastructure and tools optimized for AI workloads, are developing and training generative AI models, and have products and services that use those and other AI models.

Competition authorities have adequate tools at their disposal to deal with vertical integration in case they identify a concern. In particular, merger control and rules against abuses of dominance have been successfully used many times in the past to address potential anti-competitive effects resulting from vertical integration.

What is the rationale of the investments and/or acquisitions of large companies in small providers of generative AI systems and/or components, including AI models? How will they affect competition?

As an association that operates pre-competitively, we cannot be specific about supposed rationales in this context. What we can say, however, is the following: The role of vertical integration in the value chain should be closely monitored as the market develops (compare also answer to question 9)

Do you expect the emergence of generative AI systems and/or components, including AI models, to trigger the need to adapt EU legal antitrust concepts?

The current antitrust framework and legislation together with other existing European regulatory framework with the General Data Protection Regulation, Digital Services Act, Data Act, Data Governance Act, Digital Markets Act, AI Act, Net Neutrality Regulation and the Unfair Commercial Practices Directive seems to be fit for purpose for generative AI for the time being.

The role/scope of the DMA is basically a potent tool to address potential concerns & market imbalances if need be. The Commission should do a market investigation to identify the concerns to address in order to avoid unintended consequences that would harm competition in the generative AI sector.

Do you expect the emergence of generative AI systems to trigger the need to adapt EU antitrust investigation tools and practices?

The Commission (and other antitrust authorities) should continue to examine and monitor the generative AI space. The continued build-out of DG COMP's Data Analysis and Technology unit can be useful in that respect. The Commission does not need new investigative tools or practices as the current tools are sufficiently effective to address potential concerns.

For Europe to be sovereign in the field of generative AI, it must be involved in shaping them from the outset and also make up for the omissions in digitalisation that have been made in recent years. As experience shows that companies in Europe must make do with less financial resources and human capital than in other regions of the world, it must be in Europe's interest to play an active role in developing generative AI through innovation partnerships and interoperability. The focus should therefore lie on international cooperation for openness, standards and interoperability.

Bitkom represents more than 2,200 companies from the digital economy. They generate an annual turnover of 200 billion euros in Germany and employ more than 2 million people. Among the members are 1,000 small and medium-sized businesses, over 500 start-ups and almost all global players. These companies provide services in software, IT, telecommunications or the internet, produce hardware and consumer electronics, work in digital media, create content, operate platforms or are in other ways affiliated with the digital economy. 82 percent of the members' headquarters are in Germany, 8 percent in the rest of the EU and 7 percent in the US. 3 percent are from other regions of the world. Bitkom promotes and drives the digital transformation of the German economy and advocates for citizens to participate in and benefit from digitalisation. At the heart of Bitkom's concerns are ensuring a strong European digital policy and a fully integrated digital single market, as well as making Germany a key driver of digital change in Europe and the world.

Published by
Bitkom e.V.
Albrechtstr. 10 | 10117 Berlin

Contact person
Lukas Klingholz | Head of Cloud & Artificial Intelligence
T +49 (0)30 27576-101 | l.klingholz@bitkom.org

Dr. Pablo Schumacher | Policy Officer for Digital Content & Law
T +49 1516 2393353 | p.schumacher@bitkom.org

Title image
© Kio – pexels.com