

Call for contribution competition in virtual worlds and generative AI

1. 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..... 2
2. 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..... 3
3. 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? 4
4. 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..... 5
5. How will generative AI systems and/or components, including AI models likely be monetised, and which components will likely capture most of this monetization? 5
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.
6
7. 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? 6
8. 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?
7
9. 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..... 7
10. 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?... 8
11. 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? 8
12. Do you expect the emergence of generative AI systems to trigger the need to adapt EU antitrust investigation tools and practices? 9

1. 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.

We think that - in order of importance, the main components required to consider having AI production pipelines are:

- A proper computation infrastructure and architecture. This part is not the most uncertain nor with the more technical complexity. But this kind of choice is very dimensioning in terms of future capabilities. If you choose native cloud (use of the embedded and managed service from a cloud provider), cloud agnostic (only use cloud computing server and embed all your technical stack by your own), in-house (use your own computational power inside your organization with an overlay of cloud provider infrastructure) or even investing in your own computation station, will define, and probably limit some future technical choices. Taking into account the big picture of your AI ambition is key at this point.
- A proper Data Structure. If our "Era of AI" were the Far West, data would be Gold. Before starting any AI Journey, you should consider your data ecosystem. How easily you can have access to the data you need to accomplish your vision. Is there enough amount of data to avoid too much bias, is there enough diversity and mix in in your data to avoid fitting issues (over/under) (*overfitting is where a model stick to one result whatever the data entry are under fitting is the model cannot define an output wherever the entries are*). Depending on your data type, is your architecture suited for storage, network exchange, ingestion, curation? Are your data clean enough to not spend a tremendous amount of time in data cleaning and data curation? Before starting an AI journey, you should - with a high level of priority - perform an exhaustive data review and analysis.
- A Model Selection suited for your environment. AI models and even more for Generative AI model's environment is growing very fast and the multidimensional complexity of model selection is not to be neglected, from licensing type, to model size and inference time, or efficiency in fine tuning. You have a huge amount of intricated parameters to consider before starting defining an AI/ Generative AI solution. Starting from the end user need and product requirement is always a good practice, you should select your model in accordance with the use you will make from it.
- MLops Pipelines. This Part dives in depth with AI technology. The quickest and efficient way to start with ML ops is to take advantage of cloud native system and services, you can use for instance a combination of AWS sage maker to build a pipeline of test train models that will have a direct access onto your data structure like Aurora, or S3 datalake, you can even prepare embeddings content with open search and langchain. If your aim is to have a more simplistic out of the box use of Generative AI you can go with AWS Bedrock manage service to make one click LLM endpoints deployment, that can also access Knowledge base and vectorial database, for more precise context tuning. Another way much more difficult but with more control over your data is to build from your own your custom pipelines. The use of Docker container is now a standard in the AI industry, deploying stacks of training container that nvidia provides is a must have. You can in the same way

deploy after that with CI/CD technologies, inference endpoints in the same way. The access to those same point became afterward a simple IT problem, like configuring API endpoint access.

2. 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.

Main barriers can be identified in two areas:

1/ The first one is mainly related to the data component: data access and privacy. As previously mentioned, Data is gold in AI era, and accessing a pool of relevant data to create business specific model is quite challenging. We can identify several cases:

- A private domain company who holds the IP on its own data. This case is the easiest one from this perspective. IT and Data science skills need to be ramped up in order to create internal AI solution. However, working with a scope of data coming from an own corporation can induce a lot of bias and fitting issues due to the specificity of the data.
- A tech company that does not hold data but has technical skills. In this case, the company will need to find partners who own business context data, and to go through a high level of legal and administrative work in order to use those data. Public data could also be used of course, but in this case, it will not be possible to achieve a high level of efficiency in business specific application due to the lack of context-specific data.
- A public structure such as a university or an NGO. They will probably have a slightly easier access to data due to non-business interest but will still face the same issues as a tech company who does not hold the IP onto his own data's.

2/The second one addresses the strategic choice in balance between cost efficiency and system owning. It is mainly related to infrastructure and models selection.

- A cost efficiency approach to reach rapidly a specific application implies to delegate to a third-party provider the hosting and management of your model and "consume" it like an API. In this case, you will be charged for your consumption. Big players like ChatGPT, Mistral and Anthropic provide "pay as you use" API with a fair pricing. But in this case, you lose full control over your data - you are not sure about how it is going to be handled by the provider in reality.
- A self-hosted and self-managed AI system enables better data control. However, in this case, you have to provide an infrastructure with enough computing power, which comes with associated costs. There will be no specific costs for each request but only for your

infrastructure. This case is relevant if you take the decision to go fully into Generative AI consumption, and the amount of request you will make will absorb the price of your infrastructure. This approach should also take into consideration during your development and prototyping phase, where you will implement a costly infrastructure for a while without being able to fully take advantage of its benefits. So, the quicker you will be able to develop your solution the less longer your infrastructure will run dry with high cost.

- A last point that should be considered now, besides technical aspect, is the environmental cost. .

3. 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?

The key to success goes far beyond only Generative AI, and needs to be addressed with a complete digital solution scope. User experience is a key factor to success. Successful Generative AI company would rely on:

- An efficient access to the tools and the human machine interface. Product design is here the skill to empower this lever.
- Integration and system operability, to allow a like almost all new mainstream technical solution the multidirectional communication with third part système in order to easily be integrated in the technical ecosystem
- Scaling capabilities to always ensure a reasonable time to result without too much lagging.
- Of course, strong performance results , provided by an exhaustive and diversified dataset for training.
- Continuous improvement of your model to always stay at the best efficiency level. Here a good MLOps infrastructure for periodic retrain is the best approach.
- Have a well Scoped Technical IP and proprietary technologies to not depend on third part provider when you need to upgrade or refactor your solution.
- And besides technical regards, a well targeted market and key customer to be sure to fit the solutions to their needs. In this regard, internalized Subject Matter expertise or strong co-development partnerships should make the difference.

4. 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 terms of business applications, a first competition issue will stick around the data. Generative AI model “manufacturer” is a skill that only a few can master, and the trend of business models is to make them on open license models, according to the hugging face platform that is recognized as the hub of machine learning and AI solution who shows an increasing number of models released in open source. So, if the trends become the rules, having access to models will not be any kind of an issue. On the contrary, developing custom business application will require specific data's, in a reasonable amount to perform with a high level of efficiency. Here we will face for sure an unfair competition, between those who have embraced digital revolution years ago, and already have enough well-structured data, and those who are/will enter the game. Without being pessimistic, a number of companies will not survive this step. . It's easier and faster for companies to acquire technological skills to create Digital Solution embedding Generative AI, rather than setting a digital approach to collect and structure data to prepare Generative AI digital solutions. In France, we can notice than the French Public Bank of Investment (BPI) has launched a program to support companies integrating AI technologies to increase productivity and enrich services (*IA Booster France 2030*, s. d.) .

In terms of infrastructure and foundation models, Big Player monopoly is also a key question in terms of competition. The entry price to create a viable market product that embeds Generative AI is not that cheap, and with an aggressive business model, big players who are already in place may introduce a market concentration, with very few spaces for new players. New player would focus on “niche market” in order to be able to enter the game.

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

Monetization cannot be addressed specifically to Generative AI solution but can be addressed as a digital value solution with the same diversity of business models. Without being exhaustive, we can mention for example.

AI consultancy, where capable companies in the field of generative AI will provide expertise and skills to customers to develop and implement specific and custom-made solutions.

“As a service” approach where the value stream is delivered through web served apps, or API endpoint to customer and be paid in a “pay as you use” model. This is an expanding business model for big players.

Infrastructure model, that is a mix between consultancy and as a service, where a technical provider builds a full Generative AI for the customer: in this model, monthly / yearly licensing is the most suited business model.

In a more innovative and visionary way to generate revenue stream. we could also imagine an “exchange model” where a provider delivers free access to models in exchange for other type of value - such as data labeling and curation. For example: a user who could access Generative AI in exchange of some text labelling. A computational exchange model where user gives access to its own computer power in exchange of model access could also be imagined.

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.
-

It will for sure. According to most of the common benchmarks, top player open-source Generative AI large language models already have very similar level of performance as proprietary models, despite the fact that they have been in development for a shorter time [[LLM Arena HuggingFace](#)]. The fact that they are open source allows to create emphasis on open innovation and contribution to those models. So, with any doubt in a short-term vision open-source Generative AI will be competing with proprietary models.

7. 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?
-

Data is key, and maybe the most important asset when addressing Generative AI solutions. They will define the final performance of a model and by consequence adoption by the users and the market. The most important aspects to take into account when addressing the data question are:

- Accessibility: Are the data legally accessible according to all law policy and regulation, for reuse in business application and context?.
- Data protection: Is the current way to process data in accordance with all law, policy and regulation?. This question will take much more importance when European AI Act will enter into force as a mandatory standard.
- Data explainability: Am I aware of all bias that my dataset can induce in a custom-made model, due to the lack of heterogeneity? Can I explain them and limit their impact?

8. 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?
-

Lack of interoperability is not only a risk, but a real threat for generative AI development and providers' success. Here again we should look at the big picture, Generative AI cannot be considered as a solution itself, but more like a tool or a plugin that creates value. Locking the access with a lack of interoperability would prevent the development of new use of this foundation plugin, which is not likely to happen. Moreover, Generative AI trend is looking forward to an open and interoperable mode of development. Not being aligned with those current trends would be de facto a real handicap.

9. 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.
-

Not necessarily. At this point, generative AI or not, it is not really the parameters that will define the competitive advantage. The key point relies more in the system architecture. Elaborating a solution that as an architecture that can rely only on assembling third party tool, can provide a competitive advantage by not needing to reinvent all components. If your business environment does not require a specific management for software of unknown provenance, it could be a relevant approach. For decades now, Software Engineering has established some development pattern and best practices specially created to manage complex solution build on third part assembly. As a basic example nowadays most software solution are built with the help of libraries. Developers did not reinvent the wheel and use existing publicly available components. Their main contribution is to make all those libraries interact together for a consistent flow of processing. If you replace the term library by Generative AI endpoint you could achieve the same results. Actually, it is already the case. Hugging face is already providing libraries that allow the consumption of Generative AI endpoints in cloud infrastructure. The main drawback appears in environment with a high degree of uncertainty, when your data are often changing, when your infrastructure will often be refactored, etc... in those contexts having leverage on the full pipe of value creation can be necessary, but this approach comes also with a huge amount of inconvenient, like the cost of maintaining the management of a big and complex architecture that integrate every level of value creation.

10. 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?

Large companies can have interest in investing or acquiring small providers in some cases, and not in other cases.

- The reason to not invest or acquire a small provider:
Thinking that a small provider has developed a specific model for a specific use case, and that the large company should acquire the small provider in order to acquire the specific model is kind of a myth. All new tech companies and small providers rely on disruptive generic foundation models, those are fully accessible for anyone. The specific model a provider claims to build would never exceed the performance of a custom model trained on internal data. So, the large company should rather integrate the skills to be able to create its own model or may go with consultancy rather than acquiring a small provider for its model.

- The reasons why a big company should invest and acquire a small company:
 - Time to develop: If a big company is under a real-time pressure against competitor and does not have the time to initiate skills integration, the company should acquire a turnkey solution even if the performance is not at full level of expectation.

Integrate skills on the shelf: Integrating the skill set to build AI / generative AI can be tricky when your core business is not in technologies and IT. Investing/acquiring a small provider is an efficient way to integrate a full stack of skills like a new company department instead of building it slowly from scratch.

11. 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?

Yes this is expected, current trends and insights are already showing some pattern that should catch attention:

- The market concentration: For now, there are only few players with the ability to create performing Generative AI models that can widely reach the market (Sadare et al., 2023) even if the way they access data is questionable on the ethical side, the matter of fact is that they are now leading the market. The upcoming regulation, notably the EU AI Act, would not allow further behavior that happened in the past for those market players. Combining the lead and advantage they have taken with a more constraining data access regulation can set a lock in the market where new players will not be able to enter.

This lock could be emphasized by the network effect, if big players like GPT keep gathering more and more users by an aggressive business model and a partial free access, no user will ultimately

try to investigate different providers. Antitrust should at least take into account the historical way big players got to their solution and maybe give more flexibility in data access for new players.

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

According to the previous response, if the concepts of antitrust themselves had to change the tools, we would expect so. Tooling is just a way to achieve a vision. If the vision changes the tools must do so.

References:

IA Booster France 2030. (s. d.).

<https://www.bpifrance.fr/catalogue-offres/ia-booster-france-2030>

Sadare, O., Melvin, T., Harvey, H., Vollebregt, E., & Gilbert, S. (2023). Can Apple and Google continue as health app gatekeepers as well as distributors and developers? *Npj Digital Medicine*, 6(1), 1-7. <https://doi.org/10.1038/s41746-023-00754-6>