Schibsted

March 11, 2024

European Commission's Public Consultation on Competition on Generative AI Platforms – Input from Schibsted

Summary

- Schibsted welcomes the European Commission's initiative to investigate competition issues related to Generative AI.
- The consultation is held at a very early stage in generative AI developments and therefore there is still much uncertainty regarding how generative AI will develop.
- Access to GPUs, computational power, energy supply, the right competence and large amounts of high quality data for training requires extensive financial resources. Access to such resources can create bottlenecks and lead to entrenched market positions. The Commission should be attentive to ensuring fair access to such inputs, to excessive concentration in the industry, as well as to contractual practices that may lead to the lock-in of customers with existing large players.
- The EU appears well equipped to deal with these issues under existing regulatory frameworks, such as merger control, competition law (Article 101 and 102 TFEU) and the Digital Markets Act, so we do not see the need for new regulation at this point in time. However, the question remains whether these tools will be able to capture unilateral unfair conduct if generative AI turns into a tight oligopoly dominated by a few firms.
- Under the current regulatory frameworks, we think the EC has the necessary investigative tools and decision making powers to handle potential problems, provided they are detected sufficiently early. To this end, the EU must ensure that it develops as soon as possible sufficient technical expertise.

About Schibsted

Schibsted is founded on a long tradition of independent news, trusted marketplaces and digital consumer services. Trustworthiness and quality are core to what we do, and when using new tools such as AI we are committed to ensuring that our implementation and experimentation uphold these ideals.

At Schibsted we are leveraging the power of AI to build the best possible digital products and services for our users and to support our employees. We are currently working on AI across the company in various ways. Schibsted uses AI for many different purposes, for instance to provide different media formats, enabling privacy-friendly advertising and optimise our distribution operations, as well as predicting how many newspapers we should print to minimise our environmental footprint.

Additionally Schibsted actively participates in research collaborations to support the development of independent Nordic AI models, being one of several industrial partners of the Norwegian Research Center for AI Innovation (NorwAI) at NTNU (the Norwegian University of Science and Technology) in Trondheim. It contributes with both competence and data. One of the big projects at NorwAI is to build a generative language model for the Norwegian language, reflecting Nordic values.

To read more about our thoughts on AI and use cases:
https://futurereport.schibsted.com/downloads/schibsted_future_report_2024.pdf

Schibsted is dedicated to promoting the responsible application of AI across and beyond our organisation, and we believe that a key part of this is to be transparent about how and why we use these new technologies.

AI has great potential for a group like Schibsted, but as we have learned through research conducted in and beyond Schibsted in recent years, there are substantial risks associated with using these technologies, both for society and for users. They can relate to issues such as misinformation and disinformation, human bias being encoded into AI systems or outcomes that are hard to explain or understand. To meet these challenges, we HAVE implemented a framework for AI risk analysis. We call it the FAST framework, and it provides a common basis for approaching risk in the areas of Fairness, Accountability, Sustainability and Transparency across the Schibsted group. The framework aims to provide support for brands, and functions across Schibsted's diverse ecosystem in identifying, managing and sharing risk in order to build and use the best possible AI-powered products and services.

Input from Schibsted on the questions posed by the EU Commission

Questions 1 and 2: What are the main components (i.e., inputs) necessary to build, train, deploy and distribute generative AI systems? What are the main barriers to entry?

Computational power: Training generative AI models requires significant computational power. This involves the use of GPUs (Graphics Processing Units) or TPUs (Tensor

Processing Units) which can process large datasets and complex algorithms much faster than traditional CPUs. These resources are crucial for processing the massive amounts of data involved in training and for performing the complex calculations required for model optimization. Effective storage and networks are also components that are required to allow the GPUs to work efficiently. Also, in a cloud-based environment we need cost-effective access to hardware and infrastructure, and are reliant on the large cloud providers. We also need algorithms/Neural Network Architecture, related to the design of the model, like Transformers for GPT models.

The substantial computational power required for training sophisticated generative AI models is perhaps the most significant barrier to entry. Access to a cluster GPU can be prohibitively expensive and for the most part unattainable at this point (for very large models that require weeks of training like a language model). Access to hardware is becoming scarce and the big companies are acquiring most of what is available. In the short to medium term, access to this hardware, either directly or indirectly, is a major deciding factor for time-to market and ability to run solutions at scale. Perhaps the biggest problem is that there is mainly one provider of GPUs at the moment (Nvidia). This barrier limits the ability of smaller organisations to compete with larger entities that have the financial resources to invest in state-of-the-art computational infrastructure.

High-speed network infrastructure: Network components are critical for facilitating the large-scale transfer of data, both within the AI development environment (e.g., between storage and compute nodes) and in deploying AI solutions to end-users. The main challenges include the initial investment in and access to high-quality network equipment.

Electric Power: Essential for generative AI's lifecycle, electric power fuels the intensive computational demands of training, deploying, and distributing AI models. Reliable power supports the operation of GPUs and TPUs, crucial for processing large datasets and complex algorithms. As AI development consumes significant energy, especially in training, access to sustainable power sources becomes vital for minimising the environmental impact and ensuring the long-term sustainability of AI technologies. Balancing AI development with access to sustainable power introduces notable barriers to entry, primarily due to the high costs and energy demands of training complex AI models. The higher expenses associated with using renewable energy sources can be prohibitive for startups and smaller entities compared to larger organisations. Additionally, the availability of renewable energy varies significantly by location, adding friction for companies in regions with limited green infrastructure. Regulatory and market uncertainties surrounding renewable energy further complicate strategic planning and investments, making it challenging for new entrants to commit to sustainable practices. These factors collectively heighten the entry threshold, especially for those prioritising environmental responsibility in their operations.

Training data: For generative AI, large and diverse datasets are essential to train models to produce new content. The quality and quantity of the training data directly impact the model's ability to learn and generate outputs that accurately reflect the input patterns. Data must also be carefully curated and processed to avoid biases and ensure that the generated content is ethical and fair. Software tools, such as ML Frameworks like TensorFlow, PyTorch, etc. that provide important libraries for building and training AI models are important components.

Human Expertise/Competence: Data scientists and engineers are some of the profiles required for designing, building and maintaining these systems. There is in general a shortage of highly qualified competence in these fields within the EU. In addition, it will also be important to upskill/reskill/rehire Product and UX expertise that are used to work with new AI tools. The challenge for European companies here is to compete with the big tech actors on salary. We also need the right human expertise to safeguard that the AI development is in line with ethical and legal standards, to ensure that AI operates responsibly.

Question 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 success of AI models is highly correlated with access to the input described in response to questions 1 and 2 above. For example, access to large, diverse, and high-quality datasets enables more effective AI training, leading to superior model performance. Likewise, having robust computational power, including access to GPUs and TPUs, allows for faster and more efficient model training and iteration. Even if the cost of some of the input is decreasing, having access to sufficient resources to scale up obviously requires large financial resources. Having access to human resources is also a key factor, i.e. the availability of a team with deep expertise in AI, machine learning, and domain-specific knowledge can develop innovative solutions and improve AI systems continuously.

Two additional soft factors may influence competition between AI models:

- Regulatory Compliance and Ethics: Adhering to privacy, data protection laws, and ethical guidelines builds trust and may open up more opportunities with end-users that are particularly sensitive to these issues.
- Sustainability: Commitment to reducing the environmental impact of AI operations can enhance brand reputation and align with consumer values.

It is important that access to inputs mentioned in response to questions 1 and 2 remains open. In this respect, there is a need for careful oversight regarding the acquisition activities of the major players to avoid excessive concentration within the hands of a few large players. The risk is otherwise to see the emergence of a very small number of walled gardens, effectively strangling access to solutions not directly or indirectly sold through the big three (e.g Google, Amazon and Microsoft).

Question 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.

Some of the competition issues that we can foresee are:

- Market dominance by large players like Google, Microsoft, Meta, etc. These companies have access to enormous financial resources, large amounts of data, highly skilled people and large clusters of processing power.

- Data monopolisation: this could also be an issue if we get to a point where few companies control the majority of data that is valuable.

- Availability of talent/skilled workers is also very important, since large companies will acquire most of the talent and this could lead to a reinforcement of their market position. European institutions and national governments must do their utmost to train the workforce to ensure availability of talents in the long run.

- Platform dependency could also be an issue, since gen AI is dependent on a range of software tools and hosting systems that are mostly provided by cloud providers like AWS, Google and Microsoft. Such a high level of concentration can give rise to parallel (high) pricing, unfair terms and conditions, lock-in effects, interoperability and data portability issues for business users. The fact that these companies are also developing and commercialising AI models may also give rise to classic competition issues related to vertical integration.

- Non-compete clauses and other contractual restrictions: it is important that the rolling out of the currently available large closed AI models does not lead to significant lock-in effects through contractual clauses. For example, as a general rule, the output of in-licensed AI models should be owned by the user of the model (not the technology provider), and the user should be free to use the output of the model for any purpose, including developing, fine-tuning and commercialising alternative AI models. By definition this issue is limited to closed AI models and not open source models.

- The gatekeeper dynamics observed in today's App stores could be replicated in the AI environment with large players (Google, Facebook, AWS, OpenAI) establishing new interfaces (e.g. GPT stores) that benefit greatly from feedback loops from machine and human interactions.

Question 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 of generative AI is taking place at different level of the supply chain and in different ways:

- **Chip Manufacturing:** Manufacturing specialised chips (GPUs, TPUs) for AI processing presents a direct monetization avenue, with high demand from AI developers.

- **Infrastructure Provision:** Providing the necessary infrastructure for AI operations, from data centres to cloud services, captures significant monetization due to the extensive computational resources required by AI systems.

- **Custom Solutions and Expertise:** The combination of custom AI development with consultancy offers significant monetization, driven by the demand for specialised AI expertise and tailored solutions.

- **Platforms and Tools:** Platforms that simplify creating and deploying AI models capture notable monetization by democratising access to AI.

- **Licensing out of AI models**: AI model providers are licensing out their models, e.g. through enterprise agreements allowing businesses to build, train and develop their own (often specialised) AI models and integrate them in various operational processes (e.g., CRM, Inventory management and logistics, Bookkeeping and accounting, legal, etc). Monetization typically takes place on a subscription per user basis.

- **AI as a Service**: access to AI capabilities provided as a service for businesses that do not have the infrastructure or expertise to build, train and deploy their own models, for similar use cases as mentioned in the previous point.

- **Data monetization**: access to high quality data of specific purpose, for example: art, voices, journalism, among others. See for example the commercial agreement between Open AI and Axel Springer.

- **Generation of content**: specialised content delivered quickly and at a low cost by experts in specific categories like advertising.

- **Training and Education services** for individuals and companies.

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.

Proprietary models often achieve large, step-change improvements in performance, benefiting from significant investments, focused research agendas, and access to vast, proprietary datasets by large tech companies. These leaps can extend the performance gap suddenly, setting new industry benchmarks (such as OpenAI GPT-4 and Google Gemini Ultra).

Conversely, open-source models tend to make more incremental performance gains. The open-source community's collaborative approach fosters steady progress and innovation, but it often lacks the concentrated resources and data access that fuel the rapid advances seen in proprietary models. This difference in pace and scale of innovation highlights that open-source models are playing catch-up to the breakthroughs achieved by proprietary systems.

The tooling for provisioning and deploying models (including open-source) has improved, and became more mature and accessible (Huggingface, Google Vertex, Microsoft AI Studio, etc). This evolution in tooling includes more user-friendly interfaces, streamlined integration processes, and robust support for scaling, making it easier for developers to implement and

deploy open-source AI solutions in a wide range of applications. As a result, open-source models are not only becoming more competitive in terms of performance but also in terms of ease of use and deployment, challenging the traditional dominance of proprietary models in the market. This is not dissimilar from past patterns where solution spaces initially have been highly dominated by commercial actors but subsequently became more and more populated by open source solutions. The difference here is the lack of broad and easy access to commodity resources to develop, test and operate the open source solutions in question.

Moreover, the landscape of AI research and development is experiencing a shift towards increased secrecy among large proprietary labs. These entities have become more guarded, publishing their research findings less frequently, which contrasts with the open-source ethos of sharing knowledge and tools. This trend towards opacity not only impacts the flow of information and innovation across the broader AI community but also sharpens the divide between open-source initiatives and proprietary development efforts, influencing the competitive dynamics and accessibility of state-of-the-art AI technologies. At the moment, the largest companies like Open AI and Google will likely maintain a competitive advantage with their closed models for general purposes for a foreseeable future.

Closed research is not necessarily detrimental to competition. When carried out by smaller players, it may facilitate the emergence of alternatives to large players, who can otherwise benefit from open research to improve their closed models.

Question 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, i.e. high quality data at scale, is very important for AI models including generative AI systems, as mentioned under question 1 above.

Data from media sources is typically highly curated and includes a diverse content, which is extremely useful for training AI systems. In this respect, media companies and publishers have been raising awareness of the fact that the training on their content, done via scraping of their sites by the AI-providers, should be done with due consideration to intellectual property rights. The AI Act did not handle this situation and therefore guidelines around scraping is needed. Schibsted does not think that the Copyright Directive needs to be reopened, but future scenarios relating to use of copyright protected content might be something that an AI competition framework needs to take into account.

In addition to using media content for training, there is also the aspect of extracting the value of it. A competition framework should allow media companies and publishers to protect their intellectual rights so they can invest in creating, internally, the necessary AI competences and create value from their archives.

Fairness concept, such as the one used in the Digital Markets Act, could be used to regulate the interactions between large providers of AI models and data holders, in particular on the use of data by the former. For example, owners of key training data, such as news data, should be fairly compensated for the use of their data. This could involve monetary compensation, or licence rights to the trained AI model at advantageous terms.

Because data is such a valuable resource for AI models, it is important to ensure its availability. Output of a given AI model should be available to train other models, otherwise there is a risk of building entrenched positions.

Question 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?

We consider interoperability of generative AI systems and models relevant for competition. In particular the open interactions between generative AI systems and internal data assets is relevant for a more competitive service offer to consumers.

However, any interoperability regulation should protect intellectual property rights for all market participants. This ties to the previous question, for example, generative AI models should have an open API to fine-tune the system for specific purposes using data providers (e.g. media companies)'s own content. However, generative AI models should not be able to use media content without a previous agreement.

Also, interoperability among AI models would be an advantage for adoption and will also accelerate the evolution of this technology. Both aspects impact competition in this space and bring benefits to the market. Preserving the possibility for business users licensing in AI models to use output across models is also important to avoid lock-in effects.

Question 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.

Given the importance of certain inputs to develop AI models (see answers to questions 1 and 2), it is fundamental to guarantee open access to computer hardware and cloud platforms services at fair conditions. Absent this, vertically integrated firms may enjoy decisive competitive advantage and be able to foreclose non-vertically integrated competitors. This scenario is similar to the telecom sectors, access to networks should be possible for different services providers in order to allow competition.

Small providers of generative AI systems or components most likely do not have the resources to deploy the hardware needed for processing, so they need to rely on cloud platforms and access to them should be guaranteed at fair terms.

Moving up in the value chain, access to foundational models and interoperability should also be considered in a competition framework via standard APIs to nurture competition in the top layers/downstream markets (applications and services).

We see the possibility of vertically integrated companies developing/acquiring competitive advantages on applications and services. On the one hand generative AI is often

provisioned on the open internet protocols and is not limited/locked into a hardware device (in contrast with, for example, iPhone or Android). On the other hand, a generative AI company excelling in UX and model performance and becoming the de facto entry point to the internet may give significant competitive advantages, mirroring Google's dominance in the search domain.

Here's how this could unfold:

- **User Experience as a Differentiator:** Exceptional UX can make a platform or service much more appealing to users, encouraging adoption and habitual use. If a company manages to create an interface or ecosystem that significantly simplifies or enhances online interactions, it could become the preferred gateway for internet access, much like Google has done with search.

- **Data Network Effects:** The company would have access to vast amounts of user data, enabling it to refine and personalise the user experience and underlying AI models further. This creates a positive feedback loop, where improved services attract more users, generating more data, which in turn is used to enhance the services even more.

- **Ecosystem Lock-in:** Users accustomed to a seamless and integrated experience may find it inconvenient to switch to alternative platforms or services. This lock-in effect strengthens the company's market position, as users are more likely to explore and adopt other services offered within the same ecosystem, leading to issues of leveraging and self-preferencing.

- **Advertising and Monetization:** Dominance as an entry point to the internet places the company in an enviable position for advertising and monetization, similar to Google. The platform could attract a significant share of online advertising spend, leveraging its user base and data insights for targeted advertising and marketing services.

- **Barrier to Entry:** Achieving this level of integration and user satisfaction sets a high barrier for competitors. New entrants would need to offer substantially better or different value propositions to lure users away from the established ecosystem.

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?

Similarly to what has been seen in other industries (like pharmaceuticals), it seems that one rationale for large companies acquiring small providers of generative AI systems is to compensate for a deficit of in house innovation and capture new opportunities through external growth. This is obviously rendered possible by the enormous financial resources of large tech companies and their ability to pay acquisition prices/take risks that traditional financiers are not ready to pay/take.

Question 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?

EU legal antitrust concepts are adapted to capture several issues we anticipate in relation to generative AI. For example, merger control allows regulators to review and prevent anticompetitive mergers and acquisitions. Article 101 and 102 can be used to catch agreements and unilateral conduct by dominant firms. Antitrust rules may however not be fully adapted to regulate unfair unilateral conduct by strong yet non-dominant actors, which may become an issue in an environment dominated by a few strong players (tight oligopoly). It remains to be seen whether the Digital Markets Act is flexible enough to be applied to generative AI related issues. Otherwise, in the longer term, regulation may be necessary to unlock some markets or ensure/restore fairness and contestability.

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

Not at this stage. The European Commission is equipped with tools and powers which appear sufficient to investigate and remedy anticompetitive conduct, for example requests for information, and decisional powers to impose fines or structural remedies. Importantly, beyond tools and practices, the European Commission must ensure that it has the appropriate resources, in particular technical expertise, to analyse and understand generative AI concepts and issues. The recruitment of a CTO (and associated team) is a step in the right direction.