AI and Competition:
Breaking Barriers for Startups

Date: November 2024

Bundesverband Deutsche Startups e.V.
Schiffbauerdamm 40
10117 Berlin
Tel.: +49 (0) 30 65 77 14 34
politik@startupverband.de
www.startupverband.de

Background

Artificial Intelligence is a great opportunity for European startups. In order to create a level playing field, there is not only one single measure to take but a wide-ranging set of measures are needed. The reason for this is that Big Tech companies use multifaceted tactics to leverage their existing market power, but also that there are various markets with different needs involved.

The most serious bottlenecks European startups currently face when it comes to using and scaling AI are: **access to (high-quality) data**, **dependency on infrastructure** in the hands of Big Tech companies, and **access to talent**. While the latter can probably not be addressed by means of competition law, the others can.

Making AI work for European startups – 3 specific measures

1. **Formally designate gatekeepers' cloud services as core platform services (CPS) under the DMA and include LLMs as an additional category of possible CPS to then designate them as such.**

This would ensure that services like Microsoft Azure, Google Cloud, or AWS will be designated as Core Platforms Services, with the stringent obligations that come attached to it, in particular the ban on self-preferencing.

Even if they do not meet the quantitative thresholds needed to be designated as such under the DMA, the Commission should nonetheless carry out a market investigation of cloud services, which we believe should be concluded with their designation as CPS based on qualitative criteria, relying on Art. 3(8) and 3(9) DMA[1].

At the same time, the Commission should include LLMs as a new category of CPSs through Art.17 DMA[2], in order to designate services such as ChatGPT as CPSs. That can also be done based on a qualitative assessment, should the quantitative threshold not be met. Although formally developed by OpenAI, in fact, ChatGPT

[1] Apple's iPadOS designation as gatekeeper after the Commission's market investigation is a valuable and undeniable precedent which can be use to do so for Cloud services

[2] Which expressly states that " *The Commission may conduct a market investigation with the purpose of examining whether one or more services within the digital sector should be added to the list of core platform services or to detect types of practices that may limit the contestability of core platform services or may be unfair and which are not effectively addressed by this Regulation.* "

should be deemed a Microsoft service due to Microsoft's multibillion dollar investment in the company and increasing role in Open AI's management.

Only then, DMA rules including the ban on self-preferencing of Art 6(5) DMA, would extend to the core of AI infrastructure. This is crucial to prevent dominant players in these fields from leveraging their dominance in one AI-related market such as cloud infrastructure and to envelop more downstream AI markets. The ban on self-preferencing would then cover the preferential treatment of Gatekeepers' own foundation models on their cloud infrastructure as well as a preferential treatment of their own specialized AI services for customers who already use the gatekeeper's foundation model.

2. **Grant business users access to data that only Big Tech has and that is crucial for developing AI models.**

Large tech companies like Microsoft, Alphabet, and Amazon have access to proprietary data from their vast user bases. This data includes personal preferences, user behaviors, and real-time inputs like search history and geolocation data, which are invaluable for developing AI systems that can offer personalized and responsive services. Without access to these data, there will never be a level playing field for the development of specialized AI products.

a. **Data insights into aggregated customer behavior.**

In many industries, rich aggregated data would allow platforms to offer personalized product recommendations and improve user experiences. The idea is to predict the future behavior and needs of an individual based on patterns derived from historical data from many individuals.

Smaller companies without access to such data are at a significant disadvantage. Having limited access to proprietary data, smaller players are forced to harness the data themselves which is much less effective and accurate. This constraints their ability to develop AI models which results in less accurate recommendation systems and lower personalization.

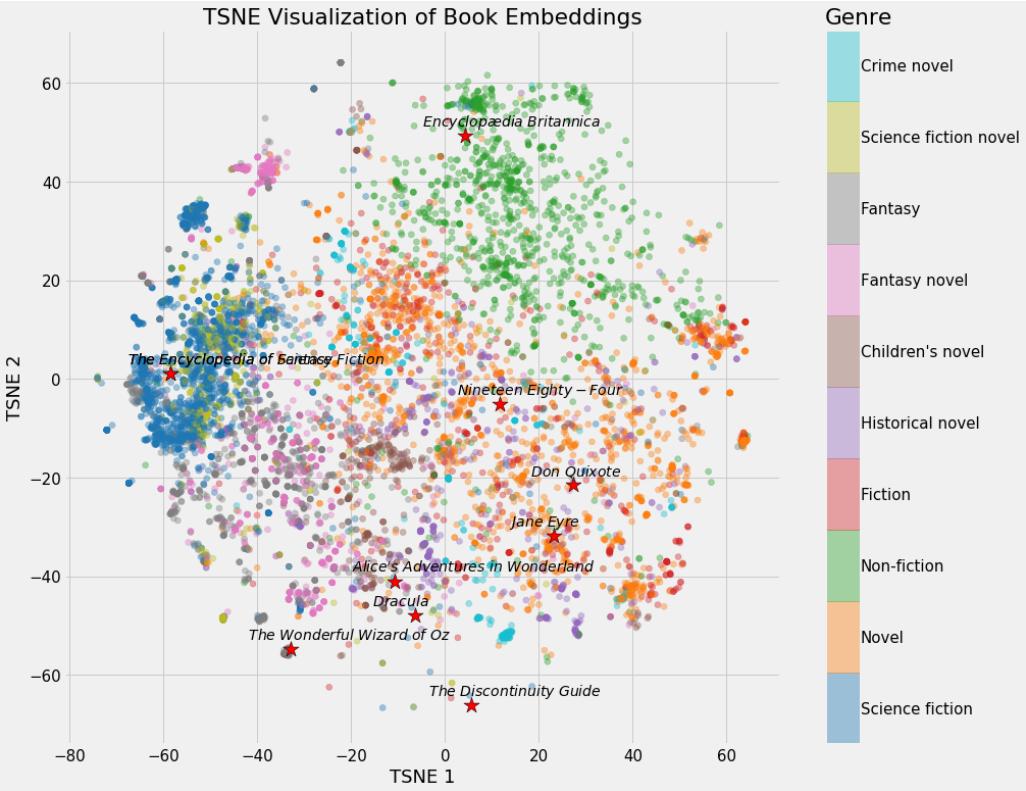
One of the biggest data aggregators is Alphabet, owning extraordinary amounts of data from a variety of sources like search history, visits of third-party websites,

Google Maps histories, and Gmail. These data enable Alphabet to extract relevant patterns of likely customer behavior for any industry. For example: Alphabet can predict with accuracy where travelers are going, when they are going, what they want to buy and with which purchasing power.

At the moment, Alphabet only shares these data insights in a way that ensures no competitor can actually use them. This is a mere PR stunt. Alphabet should be obliged to share all of these insights on a more granular basis, so that they can be used for developing specialized AI tools.

These data can be easily shared by using **embeddings**. Embeddings are aggregated representations of data in a mathematical form. They make it possible for computers to understand the strength of the relationship between words and concepts. Practically, embeddings can be used to determine the proximity of a user (represented by its aggregated data) and an offering - for instance a product the user might be interested in buying. Calculating this proximity is foundational to recommending products and services to customers.

Example: The image below could show a hypothetical book e-commerce shop. This representation could be the shop's catalog. User preferences for certain books could be characterized as the proximity to the different topical clusters[3].



[3] Picture Source: Stack Exchange, 2020

b. Live data on consumers coming through Google Ads.

In addition to aggregated data, companies using AI would also need to have access to live data on specific customers coming through Google Ads. It is crucial to understand that Google's ad customers are at the same time Google's competitors in a variety of different markets. These competitors need to use Google search as a gateway to their customers so Google controls which data these competitors get to see and which not. Right now only Google can use most of the data they have for improving their own search and discovery functionalities in numerous different services. Data used by Google for personalizing their search functionalities and product offerings in various markets include the following data types:

- **Behavioral Data**: First and foremost Google uses the **search query**, so the exact words the user typed in on Google Search. This is at the same time the most important information for Google's ad customers when it comes to leveraging AI to provide a great customer journey. Google shares the search query with advertisers using traditional ad formats but not with some of their newer ad formats. Google deliberately took the decision to keep this data point for themselves in the future. Additionally, Google uses information about how users interact with first party and third party websites and apps, such as pages visited, time spent on each page, and actions taken (e.g., clicks, downloads).
- **Transactional Data**: Purchase history and other transaction-related information related to Google's products and third party products that can help Google understand consumer preferences and buying patterns.
- **Demographic Data**: Basic information like age, gender, location, and language preferences.
- **Engagement Data**: Data from interactions with emails, social media, and other communication channels.

It is understood that there are data privacy concerns around sharing all of these data with advertisers. However, we believe that Google's competitors need a level playing field. **Should it be illegal under the GDPR to share some of the above**

mentioned data types with advertisers, Google should be forced to ignore these data types for personalizing their own services, too.

A reasonable way to grant startups access to some of these data is to **broaden the scope of the ongoing non-compliance investigation** against Google under Article 6(5) DMA. The investigation was initiated due to Google's unlawful self-preferencing within Google Search[4], and it should be revisited by issuing a **Supplemental Statement of Objections** (SSO) regarding Google's refusal to share Google Ads' data to third-parties.

Accordingly, the Commission should **expand the scope of the current investigation to include Article 6(10) DMA**. Article 6(10) requires gatekeepers to provide business users[5] with access to data generated through their interactions with end-users[6] when using the gatekeepers' CPS.

3. **Intervene to avoid lock-in effects in the AI space and cloud.**

Lock-in strategies, where tech providers limit their customers' ability to switch services, remain a challenge despite regulations like the Digital Markets Act. In AI, startups currently benefit from pay-per-use access to Big Tech's large language models (LLMs), but the introduction of long-term contracts could undermine this flexibility. Lock-in effects can also be established by technical hurdles for migrating to another provider or making it easier to buy additional services from the cloud-provider itself instead of buying them from competitors. All of these tactics could stifle innovation and impose financial burdens on startups, making it harder to change providers. The European Commission and national authorities should monitor this trend closely, and act swiftly if concerns are raised by European startups.

[4] While there hasn't yet been an instance of scope extension in a DMA case, the Commission has a well-established practice of expanding inquiries if additional evidence or complaints arise. The AdSense case (2019) for instance was initially based on Google's ad placement practices, and then expanded to include Google's use of restrictive contractual clauses with third-party websites.

[5] Who, according to Article 2(21) DMA are "*natural or legal person acting in a commercial or professional capacity using core platform services for the purpose of or in the course of providing goods or services to end users*", and therefore also European AI startups.

[6] Specifically, access to data "*that is provided for or generated in the context of the use of the relevant core platform services or services provided together with, or in support of, the relevant core platform services by those business users and the end users engaging with the products or services provided by those business users*" according to Article 6(10) DMA

Google has recently filed an antitrust complaint against Microsoft in the EU, alleging that Microsoft leverages its dominant position in business software to tie clients to its Azure cloud services. Google claims Microsoft imposes punitive fees for customers attempting to transfer projects from Azure to competing cloud providers, costing European businesses approximately 1 billion euros annually. This complaint may prompt a formal inquiry by European competition authorities, adding to ongoing investigations into Microsoft's cloud practices by the UK's Competition and Markets Authority and a separate EU probe into its bundling of Teams with other business products.

In cloud services, while Alphabet and AWS allow free migration, Microsoft Azure, the largest global cloud provider, enforces stricter lock-ins[7]. The Commission should then prevent cloud services from demanding excessive fees for the outflow of data while making data inflow available free of charge. Discount tactics that seek to tie companies to one provider for longer periods should also be restricted, while strict interoperability prescriptions should become compulsory. Only then, competition can be successfully promoted and innovation unleashed.

[7] As shown not only by Alphabet's recently filed lawsuit verting on said practices, but also by CISPE, underlining the hidden yet substantial cost behind Microsoft's anti competitive practices in cloud.

About the
German Startup Association

The German Startup Association (Bundesverband Deutsche Startups e.V. – "Startup-Verband") is the voice of startups in Germany. It represents their interests to politics, the economy, and the public. With a network of 1,200 members, the association fosters exchange among startups, scaleups, investors, and established businesses. The goal of the German Startup Association is to make Germany and Europe more startup-friendly locations.