

Carugati, Christophe

Working Paper

Antitrust issues raised by answer engines

Bruegel Working Paper, No. 07/2023

Provided in Cooperation with:

Bruegel, Brussels

Suggested Citation: Carugati, Christophe (2023) : Antitrust issues raised by answer engines, Bruegel Working Paper, No. 07/2023, Bruegel, Brussels

This Version is available at:

<https://hdl.handle.net/10419/274213>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

ANTITRUST ISSUES RAISED BY ANSWER ENGINES

CHRISTOPHE CARUGATI

Rapid development of generative artificial intelligence chatbots like ChatGPT is leading search engine providers to move from search to answer engines. Unlike search engines, which provide search results in the form of blue links to content creators, answer engines generate personalised answers through a conversation with end users. This revolution impacts the internet ecosystem of content creators and the digital advertising market. This paper outlines some early antitrust issues related to answer engines, from the transition from search to answer engines (sections 2 and 3) and the response competition authorities should adopt (section 4). It finds that search and answer engines complement and compete with each other. While the answer-engine market is still at an early stage of development, it already raises some competition issues in relation to data scraping, vertical integration and unfair terms and conditions. Intervention by competition authorities is more likely than not to prevent market power in this new market. In this regard, competition authorities should act to preserve dynamic competition and minimise adverse effects on content creators. Finally, the paper concludes with several research questions for future research (section 5).

JEL Classification: K21, L1, L4, L5, L86

Keywords: digital competition, antitrust, competition law, competition law and economics, answer engine, search engine, generative AI, ChatGPT

Christophe Carugati (christophe.carugati@bruegel.org) is a Research Fellow at Bruegel.



Recommended citation:

C. Carugati (2023) 'Antitrust issues raised by answer engines', *Working Paper 07/2023*, Bruegel

1 Introduction

The ultimate search engine, “*would understand everything that you asked it, and give you back the exact right thing instantly,*” said Larry Page, the co-founder of Google, back in 2006¹. Until now, search engines have provided what end users want in the form of organic search results of lists of content creators via blue links. But the advancement in generative artificial intelligence (AI) chatbots forces search engine providers to move from being search engines to being answer engines. Unlike search engines, generative AI answer engines generate an answer through a conversation with the end user (a so-called ‘conversational answer’). Answer engines drastically diminish search costs and increase productivity for end users because they provide what users want directly without the need to click on search results. This revolutionary change from search to answer engines impacts the internet ecosystem of content creators and the digital advertising market.

Internet actors, including search engines and content creators, rely on blue links to get traffic and generate revenue. Search engine providers rank blue links and derive revenue from advertising by promoting sponsored blue links above organic search results – a market worth \$260 billion in 2022 based on worldwide search advertising spending². Some search engine providers also derive revenue from non-search advertising, including banner or video ads, by offering advertising services, such as ad inventory, to content creators (CMA, 2020). Google thus generated about \$224.5 billion in 2022 from digital advertising, including \$162.5 billion from search advertising and \$62 billion dollars from non-search advertising³. Content creators also generate revenue from digital advertising thanks, in part, to the traffic they receive from search engines. If answer engines lead to significant traffic loss to blue links, search engines and content creators will lose advertising revenue.

Nevertheless, despite the revenue loss, search engine providers cannot avoid moving from search to answer engines. They cannot afford not to compete because newcomers, such as ChatGPT, threaten ‘traditional’ search engines, forcing the latter to innovate. ChatGPT, provided by Open AI, is the leading answer engine. In January 2023, after only two months of existence, ChatGPT already reached 100

¹ Richard Wray, ‘The ultimate search engine - Google predicts the future’, *The Guardian*, 23 May 2006, <https://www.theguardian.com/media/2006/may/23/digitalmedia.googlethedia>.

² ‘Search Advertising’, *Statista* [accessed 23 February 2023], <https://www.statista.com/outlook/amo/advertising/search-advertising/worldwide>.

³ ‘Annual Revenue of Alphabet From 2017 to 2022, by Segment (in Million U.S. Dollars)’, *Statista* [accessed 23 February 2023], <https://www.statista.com/statistics/633651/alphabet-annual-global-revenue-by-segment/>.

million monthly active end users⁴. As a comparison, it took 2.5 years for Meta-owned Instagram to gain 100 million users⁵. The move to answer engine is thus a matter of survival for search engines. Not competing means a risk of being displaced by new entrants in a new market.

In this context, search and answer engines are both complementary and competing. They complement each other as they support each other to run. Search engines need answer engines to survive. Conversely, answer engines need search engines to provide references to the generated answer.

They compete as they both provide what users want. It follows an intense dynamic competition for the transition from search to answer engines by incumbent search engine providers and new entrants. Nonetheless, answer engines already pose several competition issues. Indeed, answer engines collect and use data, including copyrighted data, to train their models and generate revenue. In jargon, they scrape the web, sometimes without permission, and use data as input from content creators to generate valuable content, in direct competition with content creators. They thus raise the old but still unsolved competition issue of data scraping. Answer engine providers, such as Microsoft, also integrate their answer engines into their own search engines. They promote their answer engines at the top of search results, and thus above the content of rivals. This raises the competition issue of self-preferencing when a firm promotes its own services over rivals. Some developed economies, including the European Union, have opted for a ban on self-preferencing in some circumstances. The ban will likely impact the development of answer engines. Finally, some answer engines like Microsoft rely on search-engine technology. In particular, they combine index data and generative AI technology to provide an up-to-date answer with citations. However, Microsoft is also one of the two providers of index data in the world. According to press reports, Microsoft has threatened to cut access to its index data for rivals that use its index data to develop their own answer engines⁶. If implemented, the threat would be akin to a refusal to supply input to rivals with unfair terms and conditions, thus raising the complex competition issue of refusal to supply.

The answer engine market is still nascent, with frequent innovations and market entries. At this stage, the competitive process works well. Despite some competition issues, there are not yet market power issues that would require competition authorities to intervene. Nevertheless, intervention will more

⁴ Eric Savitz, 'ChatGPT Users Topped 100 Million in January. Investors are Betting Big on AI', *Barron's*, 1 February 2023, <https://www.barrons.com/articles/chatgpt-investors-ai-microsoft-stock-51675284609>.

⁵ Ibid.

⁶ Leah Nylen and Dina Bass, 'Microsoft Threatens Data Restrictions in Rival AI Search', *Bloomberg*, 25 March 2023, <https://www.bloomberg.com/news/articles/2023-03-25/microsoft-threatens-to-restrict-bing-data-from-rival-ai-search-tools>.

likely than not be needed to prevent market power in this new market. Competition authorities should preserve the intense competition in the transition from search to answer engines. They should monitor closely that incumbent search engine providers do not impose entry barriers on new entrants in the answer engine market, in order to protect their business. They should also study the impact of answer engines on the internet ecosystem and how end users behave in relation to content creators. In particular, they should conduct behavioural economic studies on how users interact with blue links, because these are content creators' primary source of traffic and revenue. In case of adverse effects on content creators, competition authorities should monitor the compensation mechanisms offered by answer engines that send traffic back to content creators. If these compensation mechanisms are ineffective, they should consider working with stakeholders to develop effective alternatives.

The paper thus outlines some early antitrust issues related to the transition from search to answer engines (sections 2 and 3), and the response that competition authorities should adopt (section 4). While the answer engine market is still nascent and unpredictable, it raises fundamental research questions (section 5). Competition experts should study in depth the impact of answer engines on the internet ecosystem. They should also explore how large online platforms are integrating generative AI into their ecosystems, in particular in cloud and software services, and any attempts to lock-in their end users and business users into their ecosystems. Finally, competition authorities should pay attention to how the advancement of generative AI technology and legal regimes, including competition, data protection, content moderation and AI, impacts the evolution of answer engines.

2 Search and answer engines complement each other

Search engines and answer engines support each other in order to run. Search engines need answer engines (section 2.1), and answer engines need search engines (section 2.2).

2.1 Search engines need answer engines

Search engines need answer engines to survive. Search engines must compete with newcomers providing AI chatbots, such as ChatGPT. Answer engines are disruptive innovations that significantly impact the business models of search engine providers but, most importantly, how end users interact with search engines to find information. End users are no longer expecting search results but answers.

Some search engines, including Google Search and Microsoft Bing, already provide direct answers by promoting specialised results on top of the search results in an enriched format, such as in a specific box with pictures. They do so by promoting their own specialised search services, such as Google

Shopping, and by concluding agreements with third-party content creators, such as the online encyclopedia Wikipedia⁷. The advantage is that users can directly find specific information without clicking blue links. The drawback is that the search engines demote those blue links below the specialised results and display and position them in a non-enriched format, excluding rivals that provide similar specialised information. The practice has triggered several competition concerns about vertical integration. In Europe, the European Commission found in 2017 that Google had abused its dominant position by promoting in Google Search its own comparison shopping service Google Shopping over rival comparison shopping services, a practice known as self-preferencing. Still, the Commission did not dispute the rationale that Google displays its product results in a richer format to improve the quality and relevance of Google Search for the benefit of its users⁸.

However, search engines only provide answers from information gathered from a website. Unlike answer engines, they do not generate answers to what users are looking for precisely.

Therefore, search engines and answer engines provide different quality services. The former offers non-tailored and generic information, while the latter provides more specific and tailored information to the user query. Accordingly, to compete, search engine providers must offer answer engines. Even before the advent of answer engines, studies show that a direct answer module on a search engine improves user engagement on the search engine result page, reduces user effort and promotes user satisfaction (Wu *et al*, 2020). Search engine providers currently compete with answer engines either through vertical integration of their own answer engines into their own search engines, or by providing a separate service from their own search engines.

2.2 Answer engines need search engines

Answer engines need search engines to provide references to the generated answer. Answer engines use large language models (LLMs), known as ‘foundation models’, to perform human-like outputs. LLMs typically predict the next words in a sequence called ‘tokens’ (100 tokens are roughly equal to 75 words), to perform several natural language processing (NLP) tasks, from text summarisation to question answering. Answer engines, such as ChatGPT, answer queries through a conversation with end users. They require training LLMs to ensure that they perform tasks in accordance with the user’s

⁷ *Wikimedia Foundation*, ‘Wikimedia Enterprise Announces Google and Internet Archive as its First Customers; Allows New Customers to Self Sign-Up for Free Trials’, 21 June 2022, <https://wikimediafoundation.org/news/2022/06/21/wikimedia-enterprise-announces-google-and-internet-archive-first-customers/>.

⁸ AT.39740 *Google Search (Shopping)*, 27 June 2017.

intention. This is necessary as LLMs might otherwise contradict what users want, such as by harming users with harmful content. The training requires fine-tuning of LLMs. Engineers at Open AI use reinforcement learning from human feedback (RLHF) to fine-tune their model GPT-3 with human preferences so that LLMs perform helpful, truthful and harmless tasks (Ouyang *et al*, 2022).

Humans are labellers that intervene to label data and to create a database of human-written demonstrations of the desired output following prompts that they wrote and that users of the OpenAI Application Programming Interface (API) submitted. Engineers then use various AI models to fine-tune GPT-3 pre-trained language models (Ouyang *et al*, 2022). GPT-3 models require a vast amount of data, estimated at 45 terabytes of text data from several datasets that train the 175 billion trainable parameters of the GPT-3 models⁹. However, as of June 2023, the free version of ChatGPT does not provide references to generated answers. Unless end users ask ChatGPT for the source of information, there is no simple way for end users to verify the information provided by ChatGPT. This is an important issue as answer engines sometimes generate nonsensical or unfaithful answers, known as 'hallucinations', due to technical limitations of data and training models, such as heuristic bias in data collection (Ji *et al*, 2022).

References can help end users verify the information and know the source of information. Thanks to an algorithmic process, search engines typically provide those references in organic search results by ranking them from the most relevant to the least relevant. Thus, answer engines must incorporate the technologies of search engines to provide up-to-date, relevant references to the generated answer. Some answer engine providers, including Neeva (shutdown in June 2023), Microsoft, Google, and ChatGPT plus, provide references to the generated answer. For instance, Microsoft combines search results from its search engine and the Open AI's most advanced GPT models to answer questions with up-to-date information and references. This way, its answer engine provides a more relevant and up-to-date answer to the end user by citing the sources in a footnote with their URLs¹⁰.

3 Search and answer engines compete with each other

Search and answer engines also compete. The development of answer engines drives intense competition in which search engine providers and newcomers move from the provision of search

⁹ Kindra Cooper, 'OpenAI GPT-3: Everything You Need to Know', *Springboard*, 1 November 2021, <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>.

¹⁰ Jordi Ribas, 'Building the New Bing', *Microsoft Bing Blogs*, 21 February 2023, <https://blogs.bing.com/search-quality-insights/february-2023/Building-the-New-Bing>.

engines to the provision of answer engines (section 3.1). However, this transition raises several competition issues about data scraping, vertical integration and unfair terms and conditions (section 3.2).

3.1 The transition from search to answer engines

The shift from search to answer engines is inevitable. Search and answer engines both provide information to end users. While search engines offer search information mostly in the form of blue links and in an enriched format for some specialised results, answer engines generate tailored conversational answers to a user query. Answer engines thus reduce search costs and increase productivity for end users compared to search engines, provided that the answer is accurate. Therefore, answer engines are strong competitors for search engines, as they are more efficient and more relevant for users than search engines.

However, entry into the answer engine market is costly because of substantial infrastructure, development and deployment costs.

In terms of infrastructure, an answer engine provider relies on computing power, data and engineers. Providers need access to computing power to train and run LLMs. The computing and energy costs of training LLMs are substantial, being dependent on costly and energy-consuming AI infrastructure composed of supercomputers with specialised hardware that run from the cloud (Hoffmann *et al*, 2022). LLMs are trained and run on a large amount of common language data, such as English or French. That data is either proprietary data or open-source data available on community websites, such as GitHub or HuggingFace. Development of LLMs requires high-skilled engineers. Those developers are either in-house to the provider or external to it from community websites.

At the development level, an answer engine provider must use LLMs and fine-tune these models. LLMs are either closed-source models, such as Open AI-owned GPT-4 or open-source models, such as Meta-owned LLaMA. Models are pre-trained on common language data and have different sizes in terms of trainable parameters. The performance of the model and its costs increase with its size (Rae *et al*, 2021). The next step consists of fine-tuning the model on specific datasets to ensure it performs the desired tasks. Fine-tuning the model also generates substantial computing and energy costs (Hoffmann *et al*, 2023). In addition to training and fine-tuning, additional costs arise at inference time when the model delivers its output following a prompt from the user.

Finally, when deployed, an answer engine provider develops the answer engine service as a user-facing app.

Thus, it requires significant financial resources to train and run LLMs. Microsoft concluded a partnership with Open AI by investing \$1 billion in 2019 to back the development of Open AI exclusively on its cloud Microsoft Azure services. The partnership enables Open AI to access the computing power it needs and Microsoft to improve its Azure cloud capabilities in large-scale AI systems¹¹. Microsoft reiterated its support for Open AI by investing \$10 billion in 2023¹². In exchange, Microsoft expects financial and commercial returns from selling to business users cloud capabilities with access to LLMs, and from integrating Open AI technologies into its services, including search, web browsers and office suites¹³. Running an answer engine is thus much more expensive than running a search engine because of the costs of training and using LLMs. According to the Chairman of Google's parent company (Alphabet), John Hennessy, running an answer engine costs ten times more than a search engine¹⁴. However, open-source models trained on high-quality open-source datasets can significantly reduce the costs of training and using LLMs (Geng *et al*, 2023). Moreover, researchers are already developing small language models that rely on less-trainable parameters to reduce financial and environmental costs, while achieving the same performance (Schick and Schütze, 2021).

Furthermore, newcomers that combine search and answer engine technologies also need access to index and search data. Online search engine providers must crawl and index the web, i.e. discovering new content (crawling), and storing and organising content (indexing)¹⁵. The cost of indexing the web is so high that only Google and Microsoft do it¹⁶. Therefore, some competing search engines rely on Google and Microsoft index data to provide their services¹⁷. They also need access to vast search data to be relevant to users, especially for uncommon (tail) queries¹⁸. In Europe, the Digital Markets Act (DMA,

¹¹ Greg Brockman, 'Microsoft Invests In and Partners with OpenAI to Support Us Building Beneficial AGI', *Open AI*, 22 July 2019, <https://openai.com/blog/microsoft/>.

¹² Dina Bass, 'OpenAI Needs Billions to Keep ChatGPT Running. Enter Microsoft', *Bloomberg*, 26 January 2023, <https://www.bloomberg.com/news/articles/2023-01-26/microsoft-openai-investment-will-help-keep-chatgpt-online>.

¹³ Nilay Patel, 'Microsoft Thinks AI Can Beat Google at Search — CEO Satya Nadella Explains Why', *The Verge*, 8 February 2023, <https://www.theverge.com/23589994/microsoft-ceo-satya-nadella-bing-chatgpt-google-search-ai>.

¹⁴ Jeffrey Dastin, 'For Tech Giants, AI Like Bing and Bard Poses Billion-Dollar Search Problem', *Reuters*, 23 February 2023, <https://www.reuters.com/technology/tech-giants-ai-like-bing-bard-poses-billion-dollar-search-problem-2023-02-22/>.

¹⁵ AT.39740 *Google Search (Shopping)*, para 185.

¹⁶ Leah Nylen and Dina Bass, 'Microsoft Threatens Data Restrictions in Rival AI Search', *Bloomberg*, 25 March 2023, <https://www.bloomberg.com/news/articles/2023-03-25/microsoft-threatens-to-restrict-bing-data-from-rival-ai-search-tools>.

¹⁷ AT.39740 *Google Search (Shopping)*, para. 304

¹⁸ *Ibid*, paras. 287 and 288.

Regulation (EU) 2022/1925) requires online search engine providers falling within its scope, known as 'gatekeepers', to share search data with competing providers (Article 6(11) DMA)¹⁹. The provision is thus likely to lower entry costs for online search engines and, accordingly, to ease the cost of accessing search data for answer engine providers.

Still, despite these high entry barriers, current online search engine providers and newcomers frequently enter the answer engine market to take market shares from Google. Since Open AI released ChatGPT in November 2022, several firms have announced the development of answer engines, including Neeva in January 2023²⁰, Microsoft, Google and Baidu in February 2023²¹, and DuckDuckGo in March 2023²².

In this early-development stage, firms entering the answer engine market consider that the answer engine complements the search engine. This might be down to four main factors:

1. Running an answer engine for all queries is currently too expensive. Indeed, the cost of running an answer engine also depends on how many words it generates, known as inference cost. According to Morgan Stanley, if Google uses an answer engine for half of its queries to generate a 50-word answer, it could cost Google \$6 billion²³.
2. The answer engine sometimes provides inaccurate answers to end users that might lead to reputational damages for the brand. For instance, Google reportedly lost nearly \$100 billion in market value following inaccurate answers from its answer engine Bard during its first presentation to the public in February 2023²⁴.
3. The business model of answer engines is still in development. Some firms, including Microsoft and Google, are considering an advertising-based business model to replicate the business model of search engines. In this business model, end users will use the answer engine at a zero price while

¹⁹ For an explanation about the firms falling within the DMA scope, see Carugati (2023a).

²⁰ Neeva, 'Introducing NeevaAI', 6 January 2023, <https://neeva.com/blog/introducing-neevaai>.

²¹ Jennifer Jett and Larissa Gao, 'Chinese Tech Giant Baidu to Launch ChatGPT-Style AI Bot', *NBC news*, 7 February 2023, <https://www.nbcnews.com/news/world/baidu-chatgpt-ai-bot-china-rcna69453>.

²² Gabriel Weinberg, 'Duckduckgo Launches Duckassist: A New Feature that Generates Natural Language Answers to Search Queries Using Wikipedia', *DuckDuckGo Blog*, 8 March 2023, <https://spreadprivacy.com/duckassist-launch/>.

²³ Jeffrey Dastin, 'For Tech Giants, AI Like Bing and Bard Poses Billion-Dollar Search Problem', *Reuters*, 23 February 2023, <https://www.reuters.com/technology/tech-giants-ai-like-bing-bard-poses-billion-dollar-search-problem-2023-02-22/>.

²⁴ Jonathan Ponciano, 'Alphabet Stock Plunge Erases \$100 Billion After New AI Chatbot Gives Wrong Answer In Ad', *Forbes*, 8 February 2023, <https://www.forbes.com/sites/jonathanponciano/2023/02/08/alphabet-google-stock-plunge-erases-100-billion-after-new-ai-chatbot-gives-wrong-answer-in-ad/?sh=5aa34c9255ce>.

advertisers will pay for advertising in the form of paid links or promoted content in the generated answer²⁵. Other firms, including Open AI, have opted for a subscription-based business model that provides the most advanced version of the answer engine and additional services to end users that subscribe to a paid plan. Open AI also chose to generate revenue by allowing third parties to develop services by giving access to the API of its LLMs and ChatGPT for a fee²⁶. While advertising-based is an obvious business model for search engine providers, replicating such a model with an answer engine currently has a major limitation. At this early stage of development, it is too early to know how end users will behave with links or other promoted content in the generated answer. Will end users be satisfied with the generated answer, or will they click the paid links for more information? These questions are crucial as only a click to the paid link or promoted content ensures conversion and generates advertising revenue, provided that the business model relies on pay-per-click and not on pay-per-impression (views).

4. Users of search engines currently use search engines for navigational queries, namely a query to access a website (eg Booking.com) or specific information (eg weather in Brussels) and for research queries, namely a query to find general information (eg what to do in Paris for a romantic weekend). Navigational queries are unlikely to require an answer from an answer engine as they only need a direction to a website or specific information. By contrast, research queries will likely require an answer from an answer engine as they need a personalised answer. For this reason, Microsoft opted for the coexistence of the search and answer engine²⁷.

Against this background, search engine providers opt for two different approaches. All the mentioned firms, except Google, integrate the answer engine into their search engine. The integration enables them to leverage the end users and advertisers of their search engines and to ensure a frictionless transition between search and answer engines. By contrast, Google provides, for now, its answer engine Bard as a separate service from Google Search. The separation makes sense at this early stage of development to ensure that Google Search does not suffer from reputational damage arising from inaccurate generated answers. The separation also allows Google to experiment with new features and business models

²⁵ Sheila Dang, 'Exclusive: Microsoft's Bing Plans AI Ads In Early Pitch To Advertisers', *Reuters*, 17 February 2023, <https://www.reuters.com/technology/microsofts-bing-plans-ai-ads-early-pitch-advertisers-2023-02-17/>, and Vidhya Srinivasanl, 'Search Ads Today and in the Future', *Google Blog*, 10 May 2023, <https://blog.google/products/ads-commerce/google-search-ads-generative-ai/>. Google will keep its advertising model by showing sponsored content.

²⁶ *OpenAI*, 'Introducing ChatGPT Plus', 1 February 2023, <https://openai.com/blog/chatgpt-plus>.

²⁷ Jordi Ribas, 'Building the New Bing', *Microsoft Bing Blogs*, 21 February 2023, <https://blogs.bing.com/search-quality-insights/february-2023/Building-the-New-Bing>.

without compromising advertising revenue from Google Search. However, Google has already announced the integration of LLMs and its answer engine into Google Search²⁸.

3.2 Competition issues

The answer engine market is still at an early stage of development as of May 2023. Frequent market entries and innovations from search engine providers, such as Google Bard, and newcomers, such as Open AI ChatGPT, currently characterise the market. For instance, Open AI constantly releases new models (eg GPT-4) and features (eg ChatGPT plugins) for ChatGPT²⁹. Therefore, no market player, including Google, currently has market power in the answer engine market.

Nonetheless, the absence of market power issues does not mean there are no competition issues. Indeed, the answer engine market raises several competition issues regarding data scraping (section 3.2.1), vertical integration (section 3.2.2) and unfair terms and conditions (section 3.2.3).

3.2.1 Data scraping

Answer engines collect and use data from content creators to generate their answers to end users. Some content creators have already filed lawsuits in the United States and the United Kingdom against several generative AI providers which allegedly use copyrighted data without permission to generate content³⁰. The ongoing issue is essentially a matter of intellectual property rights (IP rights), and the interpretations of the issue given by national courts are likely to impact answer engine development.

While this IP rights issue is out of this paper's scope, collecting and using data without consent is also an old competition issue. It refers to data scraping, namely collecting vast amounts of data from a source, such as the web. The firm then uses data in its services or derives intelligence to develop new services. In Europe, plaintiffs in the *Google Search (Shopping)* case complained before the European Commission (the Commission) that Google abused its dominant position by scraping content from

²⁸ Sissie Hsiao and Eli Collins, 'Try Bard and Share Your Feedback', *Google Blog*, 21 March 2023, <https://blog.google/technology/ai/try-bard/>, and Elizabeth Reid, 'Supercharging Search with Generative AI', *Google Blog*, 10 May 2023, <https://blog.google/products/search/generative-ai-search/>.

²⁹ Open AI, 'GPT-4', 14 March 2023, <https://openai.com/research/gpt-4>, and OpenAI, 'ChatGPT Plugins', 23 March 2023, <https://openai.com/blog/chatgpt-plugins>.

³⁰ Blake Brittain, 'Lawsuits Accuse AI Content Creators of Misusing Copyrighted Work', *Reuters*, 17 January 2023, <https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/>. Publishers have also criticised the use of their content. See Gerry Smith, 'OpenAI Is Faulted by Media for Using Articles to Train ChatGPT', *Bloomberg*, 17 February 2023, <https://www.bloomberg.com/news/articles/2023-02-17/openai-is-faulted-by-media-for-using-articles-to-train-chatgpt>.

rivals³¹. The Commission supported the view that data scraping poses competition issues. In 2012, the former European competition Commissioner, Joaquín Almunia, considered that:

“Our second concern relates to the way Google copies content from competing vertical search services and uses it in its own offerings. Google may be copying original material from the websites of its competitors such as user reviews and using that material on its own sites without their prior authorisation. In this way they are appropriating the benefits of the investments of competitors. We are worried that this could reduce competitors’ incentives to invest in the creation of original content for the benefit of internet users. This practice may impact for instance travel sites or sites providing restaurant guides”³².

In 2015, the Commission sent a statement of objections (SO) to Google for allegedly promoting its comparison shopping service Google Shopping over rivals. The press release stated the data scraping issue was out of the scope of the SO but that the investigation was still ongoing on that front³³. In 2017, the Commission found that Google abused its dominant position because Google did not position and display competing comparison shopping services in the same format and demoted rivals by subjecting them to different algorithms³⁴. Since then, for unknown reasons, the Commission has not published any press release about the data scraping issue. The *Google Search (Shopping)* investigation thus seems closed as of April 2023.

However, the answer engine will likely reinvigorate the data scraping issue: answer engines scrape the web and use the content of content creators to generate their own content. Figure 1 shows that Microsoft uses content from various sources, including TripAdvisor, to generate an answer to a query about accommodation in Brussels. The Microsoft answer engine also prominently displays a map of accommodation in Brussels. Some content creators might partner with search engine providers to reproduce this information or tolerate the practice³⁵. Nevertheless, data scraping poses a competition issue as the answer engine generates content that competes directly with content provided by content

³¹ Fairsearch.org, ‘What to Know About the European Commission Investigation’, 23 May 2012, <https://fairsearch.org/what-to-know-about-the-european-commission-investigation/>.

³² ‘Policy Statement of VP Almunia on the Google Antitrust Investigation’, European Commission, 21 May 2012, https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_12_372.

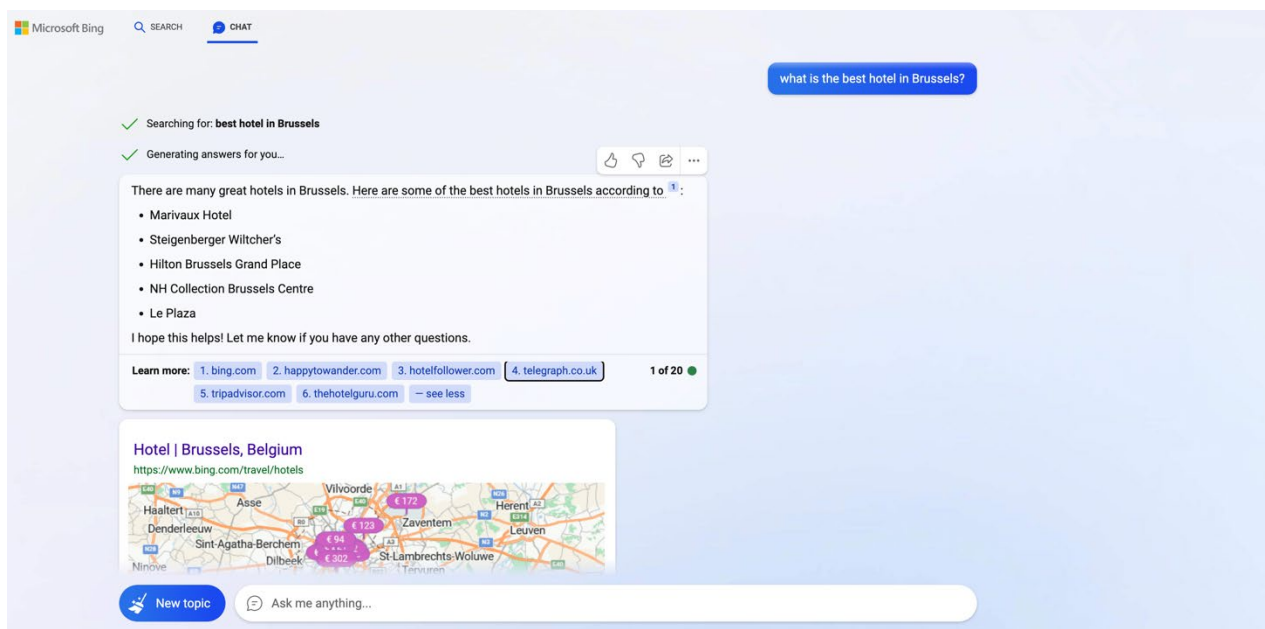
³³ European Commission memo of 15 April 2015, ‘Antitrust: Commission Sends Statement of Objections to Google on Comparison Shopping Service’, https://ec.europa.eu/commission/presscorner/detail/en/MEMO_15_4781.

³⁴ AT.39740 *Google Search (Shopping)*.

³⁵ For instance, Microsoft signed a partnership with TripAdvisor in 2013. Greg Sterling, ‘Bing Integrates TripAdvisor Tools, Content into Search Results’, *Search Engine Land*, 21 November 2013, <https://searchengineland.com/bing-integrates-tripadvisor-travel-tools-content-into-search-results-177958>.

creators on which scraping occurred. Competition law applies to this practice if scraping has a potential anticompetitive effect³⁶. This is irrespective of whether IPRs protect scraped data, as competition law applies independently from other legal regimes³⁷. The practice might exploit rivals (exploitative abuse) and exclude them (exclusionary abuse). Indeed, while there are not yet behavioural economics studies on how users will behave with links in footnotes in the generated answer, it is conceivable that users will not click on them to get more information or will click significantly less. In this case, content creators providing similar content will lose traffic and advertising revenue, excluding them from the market. Nonetheless, dominant firms could escape to a competition law infringement by proving that objective procompetitive effects, such as saving end users' search costs, outweigh anticompetitive effects³⁸.

Figure 1: Generated answer by Microsoft Bing to a query about accommodation in Brussels



Source: Bruegel from Microsoft Bing answer engine (accessed 29 March 2023).

3.2.2 Vertical integration

Some answer engine providers, such as Microsoft, integrate their answer engines into their search engines. They have thus opted for vertical integration of their answer engine. In the last few years,

³⁶ C-457/10 AstraZeneca AB and AstraZeneca plc v. European Commission, ECLI: EU: C:2012:770, 6 December 2012, para. 112.

³⁷ C-457/10 AstraZeneca AB and AstraZeneca plc v. European Commission, ECLI: EU: C:2012:770, 6 December 2012, para. 132. See also, C-377/20 Servizio Elettrico Nazionale SpA and Others v Autorità Garante della Concorrenza e del Mercato and Others, ECLI:EU:C:2021:998, Opinion of Advocate General Rantos, 9 December 2021, para. 36.

³⁸ C-209/10 Post Danmark A/S v Konkurrencerådet, ECLI:EU:C:2012:172, 27 March 2012, paras. 41-42.

competition authorities and legislators have focused their efforts on tackling a particular type of vertical integration when a dominant firm self-preferences its own services, leading to discrimination between its own services and those of competitors. In Europe, the Commission thus considered in *Google Search (Shopping)* that the practice is harmful when a dominant firm does not provide equal treatment to its services and those of rivals (Carugati, 2022). While the EU General Court confirmed the Commission's approach in 2021, the case is still pending before the European Court of Justice³⁹.

Meanwhile, worldwide, several digital competition laws and draft laws prohibit or consider prohibiting large online platforms from self-preferencing before it takes place *ex-ante* (Carugati, 2023b). In Europe, the DMA prohibits gatekeepers from promoting their services over rivals in crawling, indexing and ranking (Article 6(5) DMA). Gatekeepers cannot demonstrate procompetitive effects that outweigh anticompetitive effects. They cannot do it because the DMA does not allow for objective justification, even though some economic studies show that self-preferencing might have positive effects⁴⁰. In practice, integrating the answer engine will likely raise two issues of self-preferencing.

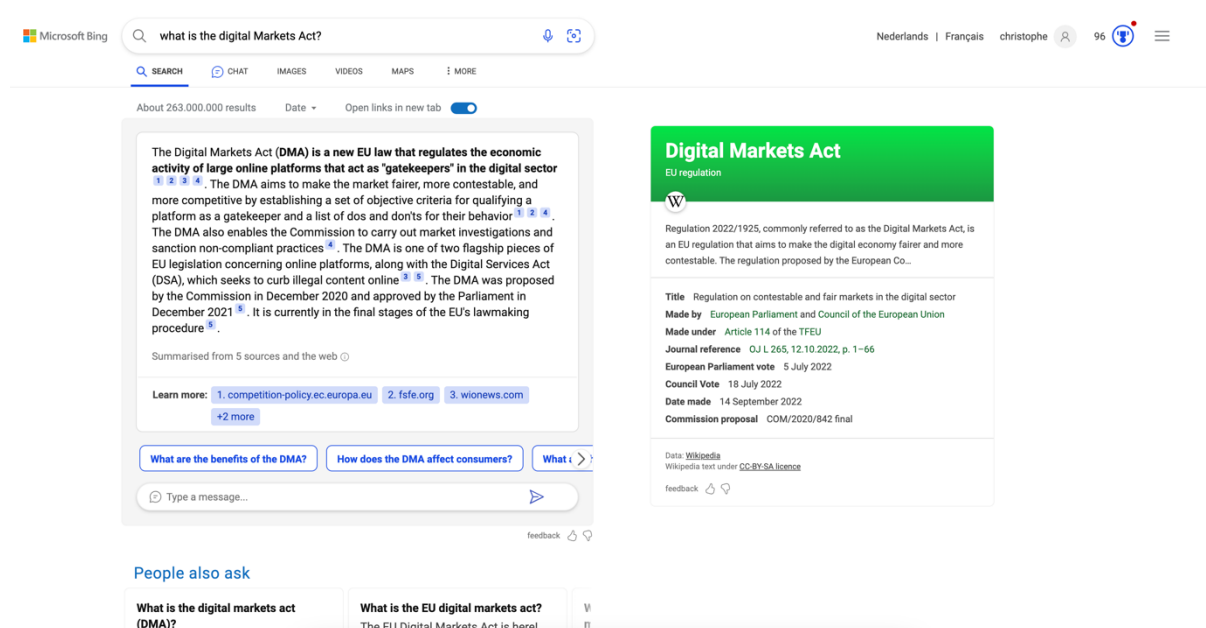
The first issue is when the answer engine provider offers its answer engine above search results and does not offer equal treatment to competing content creators that provide similar information, as illustrated in Figure 2. Content creators might argue that they lose traffic because users no longer click on blue links as the answer engine prominently provides an answer, thus eliminating the need to consult blue links below the generated answer. The practice might fall under competition law as competing content creators do not have equal treatment with the answer engine, resulting in potential exclusion from the market. Studies have already found that providing direct answers to a search query above the search results leads to shorter search sessions and fewer clicks on search results, compared to when a search engine only provides organic search results (Wu *et al*, 2020).

The second issue is when the provider offers in its answer its own specialised search service over competitors, as illustrated in Figure 1 with the map. This issue will raise similar questions as the above *Google Search (Shopping)* case and will likely be covered by the DMA prohibition of self-preferencing.

³⁹ T-612/17 *Google and Alphabet v Commission (Google Shopping)*, ECLI:EU:T:2021:763, 10 November 2021. C-48/22 P *Google and Alphabet v Commission (Google Shopping)* (the case is still pending).

⁴⁰ For a comprehensive economic literature review on the pro and anticompetitive effects of self-preferencing, see Peitz (2022). For a law and economic discussion about self-preferencing, see Bougette *et al* (2022).

Figure 2: Generated answer by Microsoft Bing about the Digital Markets Act



Source: Bruegel from Microsoft Bing answer engine (accessed 29 March 2023).

3.2.3 Unfair terms and conditions

Some search engine providers want to combine search and answer engine technologies. As noted above, several search engine providers rely on index data from Microsoft to provide their services. However, press reports highlighted in March 2023 that Microsoft threatened to restrict access to its index data if rivals did not stop using it for developing their own answer engine because, according to Microsoft, the practice violates the terms and conditions of use of its index⁴¹. Restricting access to index data would pose a competition issue. Competition law prohibits a dominant firm from refusing to supply input, such as data, to a competitor under certain conditions. The Court of Justice requires three conditions in its *Bronner* judgement. First, the refusal is likely to eliminate all competition in another market. Second, the refusal is not objectively justified. And third, the input is indispensable in a way that there is no actual or potential substitute for this input due to technical, legal or economic obstacles⁴². If IP rights protect the input, the Court of Justice clarified another condition in its *IMS Health* judgement. In that situation, a fourth condition is that the refusal prevents the emergence of a new product for which there is potential consumer demand⁴³.

⁴¹ Leah Nylen and Dina Bass, 'Microsoft Threatens Data Restrictions in Rival AI Search', *Bloomberg*, 25 March 2023.

⁴² C-7/97 *Bronner*, ECLI:EU:C:1998:569, 26 November 1998, para. 41.

⁴³ C-418/01 *IMS Health*, ECLI:EU:C:2004:257, 29 April 2004, para. 38 and para. 49.

However, in *Deutsche Telekom vs Commission* and *Slovak Telekom vs Commission*, in cases concerning telecommunication firms subject to a regulatory obligation to give access to their infrastructure, the Court of Justice considered that a constructive refusal to supply does not require proof of indispensability. A constructive refusal to supply occurs when the dominant firm provides access to the input with unfair terms and conditions⁴⁴.

If Microsoft implements its threat and an IP right does not protect Microsoft index data, the *Deutsche Telekom* and *Slovak Telekom* judgements would apply if the judgements held in a case where there is no obligation to give access under a regulatory obligation. Indeed, Microsoft already provides access to index data to rivals. However, restricting rivals from using index data in developing their own answer engines might be unfair as it would put rivals at a competitive disadvantage compared to Microsoft, which uses the index to develop its answer engine⁴⁵. The restriction would lead to anticompetitive effects as it would prevent rivals from developing competing answer engines, thus leading to their exclusion from the market. Given the Court of Justice case laws, the Commission in March 2023 updated its abuse of dominance enforcement guidance to clarify that it might investigate cases when the input is not indispensable⁴⁶. These developments are particularly relevant in data-driven markets when the proof of the indispensability of data is complex because data is sometimes replicable (eg personal data such as name and email address), ubiquitous and non-rival⁴⁷. However, if an IP right protects Microsoft index data, the *IMS Health* judgment would apply. In that case, the condition of the new product for which there is potential consumer demand would probably prevent rivals from requesting access because rivals could not argue for a new product since Microsoft already offers an answer engine. Moreover, in both situations, the condition that the refusal is likely to eliminate all competition in another market, namely the answer engine market, might also prevent firms from requesting access because there are already competitors, such as Google or Neeva, that compete with Microsoft in the answer engine market without relying on the Microsoft Index data⁴⁸. Therefore, the refusal will unlikely eliminate all competition in the answer engine market.

⁴⁴ C-152/19 P *Deutsche Telekom / Commission*, ECLI:EU:C:2021:238, 25 March 2021, para. 50 and para. 60. See also, C-165/19 P *Slovak Telekom v Commission*, ECLI:EU:C:2021:239, 25 March 2021, para. 50 and para 60.

⁴⁵ Jordi Ribas, 'Building the New Bing', *Microsoft Bing Blogs*, 21 February 2023.

⁴⁶ Amendments to the Communication From the Commission Guidance on the Commission's Enforcement Priorities in Applying Article 82 of the EC Treaty to Abusive Exclusionary Conduct by Dominant Undertakings, 27 March 2023.

⁴⁷ For a discussion of a refusal of data access and the condition of indispensability, see Crémer *et al* (2019). See also Graef *et al* (2019).

⁴⁸ Neeva only relies on the Microsoft Bing index for images; see <https://neeva.com/faq>.

4 The role of competition authorities

At this early development stage, despite the absence of market-power issues, intervention by competition authorities is more likely than not, as they want to prevent market power in this new market. In that regard, they should preserve the intense dynamic competition in the transition from search to answer engines (section 4.1). Moreover, given the impact that answer engines might have on the internet ecosystem, they should minimise adverse effects on content creators (4.2).

4.1 Preserving dynamic competition

Markets require legal intervention when there are not working well due to market failures, such as market-power issues (Paccès and Visscher, 2011). Market power leads to static and dynamic inefficiencies as it might lead to price increases, output decreases or fewer incentives to innovate. Competition authorities thus intervene to correct market power. In the absence of market power, competition authorities do not intervene as firms do not have the ability to impose harmful practices on competition and have the incentive to innovate to gain market power⁴⁹.

The answer engine market is still nascent, and no firm currently has market power. Nonetheless, dominant firms in the search engine market will try to leverage their market power in search to enter the answer engine market. As noted above, the leveraging might pose competition issues, especially when the firm envisages promoting its own answer engine over competing content creators. In that situation, competition laws and digital competition rules will likely intervene *ex-post* and *ex-ante* to prevent incumbents in the search engine market from self-preferencing their answer engines.

However, dynamic competition puts into question legal intervention. On the one hand, legal intervention would be beneficial to prevent the incumbent in the search engine market from imposing anticompetitive conditions that would inhibit new entries in the answer engine market. On the other hand, legal intervention would be harmful as the incumbent in the search engine market has a strong incentive to innovate because the answer engine threatens its position in the search engine market. It must focus its efforts on providing innovative services to its end users, as the threat comes from frequent new entrants, including newcomers that do not offer search engines, such as ChatGPT, and innovations without the possibility of anticipating the next entrant and innovation. In these circumstances, even if the incumbent can raise entry barriers to new entrants in the answer engine market, it might not have the incentive to impose them. The effects will be uncertain at best because it

⁴⁹ For a discussion of competition and innovation, see Shapiro (2011).

cannot know which barrier to impose in response to the threat, and, at worst, counterproductive by raising public backlash.

Against this background, competition authorities are unlikely to intervene as the goal of competition law is to protect the competitive process, not competitors⁵⁰. Therefore, competition authorities should refrain from intervening to avoid creating market distortions when the competitive process works, as is the case with the answer engine market. However, competition authorities are now well aware that the digital economy is fast-moving and prone to quickly tipping towards a quasi-monopoly due to the characteristics of data-driven markets, including data advantage and direct and indirect network effects, defined as when the user utility changes with the number of users in the same group (direct network effect) and another user group (indirect network effect) (Digital Competition Expert Panel, 2019)⁵¹. In particular, data-driven firms learn from data with AI to improve their offerings to users, known as data-driven network effects, defined as when the user utility changes with improved learning from data (Gregory *et al*, 2021). Answer engines also exhibit data-driven network effects as LLMs learn from conversations to improve their generated answer. In other words, the more users use the answer engine, the more the model collects and processes data to train itself, thus generating a better answer⁵². Some competition authorities are thus already sending signals to economic actors in the metaverse – an immersive virtual world – and answer engine that they are watching those markets closely to prevent domination by large digital platforms⁵³.

Intervention is thus more likely than not, as regulators want to prevent market power in new markets. In this early stage of the answer engine market development, competition authorities should focus on enabling firms to innovate and ensuring that incumbents in the search engine market do not impose anticompetitive conditions on newcomers, like those mentioned in section 3.

⁵⁰ C-8/08 *T-Mobile Netherlands BV*, *Opinion of Advocate General Kokott*, ECLI:EU: C:2009:110, February 19, 2009, para. 71.

⁵¹ In German competition law, the rule explicitly mentions data as a source of market power (para. 18(3a) German Act against Restraints of Competition (GWB)). Furthermore, expert reports and market investigations on online advertising refer to data as a source of market power. See Cabral *et al* (2021).

⁵² For instance, Open AI collects data from conversations with ChatGPT to train and improve its models. “*We don’t use data for selling our services, advertising, or building profiles of people—we use data to make our models more helpful for people. ChatGPT, for instance, improves by further training on the conversations people have with it, unless you choose to disable training.*” See <https://help.openai.com/en/articles/7730893-data-controls-faq>.

⁵³ See Vestager (2023), and, for the US, Leah Nylen, ‘FTC Is Reviewing Competition in Artificial Intelligence’, *Bloomberg*, 27 March 2023, <https://www.bloomberg.com/news/articles/2023-03-27/ftc-reviewing-competition-deception-in-artificial-intelligence>.

Moreover, competition authorities should assist the transition from search to answer engine to minimise adverse effects on content creators.

4.2 Minimising adverse effects on content creators

Answer engines pose the legitimate question about their impact on content creators that provide content to them and depend on traffic to generate revenue, mostly from non-search advertising. While answer engines might benefit end users by increasing productivity and reducing search costs, content creators might have less traffic and less incentive to produce content. Answer engines might eliminate the need for end users to consult content creators as the answer engine already generates the answers end users want, and content creators might create less content because of the loss of revenue. It follows that answer engines might lead to less advertising spending on non-search advertising if users no longer consult or consult fewer content creators. Therefore, answer engines might impact content creators and the advertising industry.

Search engine providers are aware of their impact on this ecosystem. For instance, Microsoft Bing cites the sources as, according to Microsoft, “[s]ending traffic to these sources is important for a healthy web ecosystem and remains one of our top Bing goals”⁵⁴. Open AI introduced in March 2023 a similar mechanism with its web browsing plugin that cites the source of information by stating “*this is a new method of interacting with the web, and [we] welcome feedback on additional ways to drive traffic back to sources and add to the overall health of the ecosystem*”⁵⁵. Google announced in May 2023 a similar initiative with the forthcoming integration of its answer engine into its search engine by stating that “*we’re committed to continue sending valuable traffic to sites across the web*”⁵⁶. However, as noted above, it is too early to know whether end users will click on the citations, and thus, whether this mechanism of sending back traffic to the original content creators will be effective.

In this context, competition authorities should monitor how these link mechanisms work and how effectively they send back traffic to content creators, notably by studying how users behave through behavioural studies.

⁵⁴ Jordi Ribas, ‘Building the New Bing’, *Microsoft Bing Blogs*, 21 February 2023.

⁵⁵ *OpenAI*, ChatGPT plugins, 23 March 2023.

⁵⁶ Elizabeth Reid, ‘Supercharging Search with Generative AI’, *Google Blog*, 10 May 2023.

However, some press publishers are already lobbying for a compensation mechanism when the answer engine uses their data to train its AI and when the engine does not send traffic back to them⁵⁷.

The first issue of compensation for training AI models is more case-specific and an IP right issue related to copyrighted content. This issue is thus out of the scope of this paper.

The second issue of compensation from a search engine is familiar to competition experts. In Europe, the copyright directive (Directive (EU) 2019/790, Art. 15) allows press publishers to receive compensation from search engine providers for the online use of their works (so-called 'related rights'). The French competition authority used the new copyright directive to open a case of abuse of its dominant position against Google in April 2020. The competition authority found that Google did not negotiate with and imposed unfair terms and conditions on press publishers for using their content⁵⁸. The case ended in June 2022 with commitments by Google on a compensation mechanism for press publishers⁵⁹. Whether such a related rights mechanism should be extended to all content creators will depend on the effectiveness of the above link mechanisms and other future private mechanisms that compensate content creators for using their content. In this regard, competition authorities should work with stakeholders to develop effective compensation mechanisms if they are ineffective.

5 Conclusion

The transition from search to answer engines is probably the most significant innovation and driver of competition in the digital economy. By providing conversational answers to queries, answer engines enable productivity gains and lower search costs in all economic sectors, from travel to legal. This transition profoundly disrupts the internet ecosystem. The internet is valuable because of the richness of the content that users can find online from millions of content creators. Now with answer engines, creators of content, including copyrighted content, are becoming input for a generated answer by an answer engine. Depending on how end users interact with answer engines, answer engines might quickly become the main source of content on the internet.

⁵⁷ Zoe Thomas, 'Online Publishers Want AI Chatbot Makers to Pay Up', *The Wall Street Journal*, 27 March 2023, <https://www.wsj.com/podcasts/tech-news-briefing/online-publishers-want-ai-chatbot-makers-to-pay-up/33108bd8-f9b6-4446-99d0-a0b3e06d6df1>.

⁵⁸ Decision 20-MC-01 of 9 April 2020 on Requests for Interim Measures by the Syndicat Des Éditeurs De La Presse Magazine, The Alliance De La Presse D'information Générale And Others and Agence France-Presse. Spain opened in March 2023 a similar investigation against Google: 'The CNMC Initiates Disciplinary Proceedings Against Google for Possible Anti-Competitive Practices Affecting Spanish Publishers of Press Publications and News Agencies', *Comisión Nacional de los Mercados y la Competencia*, 28 March 2023, https://www.cnmc.es/novedad/Incoacion-Google_INC_CNMC_20230328.

⁵⁹ Decision 22-D-13 of 21 June 2022 Regarding Practices Implemented In the Press Sector, 21 June 2022.

As a research question, the impact of answer engines on the internet ecosystem should be thus the priority for researchers in competition and intellectual property laws. Indeed, the roles of competition and intellectual property laws are to protect the competitive process and the content of content creators. By collecting and using data from content creators for training answer engines and generating their content over rivals, answer engine providers might negatively impact competing content creators with unknown consequences for the quality and diversity of online content.

Furthermore, while the answer engine market is currently competitive and innovative, competition experts should closely watch how the market evolves and how the integration of answer engines into the other services of large online platforms might reinforce their ecosystems and lock users into them. Press reports have already highlighted that Google, Microsoft and Amazon use generative AI as a business driver for their services, including cloud and software services⁶⁰. The president of the German competition authority stated in March 2023 that it will look at this aspect in its investigation to designate Microsoft as a firm subject to German digital competition legislation⁶¹. Finally, more broadly, researchers should pay attention to the development of generative AI and how laws in the field of competition, data protection, content moderation and AI interact and impact the evolution of answer engines.

References

Bougette, P., O. Budzinski and F. Marty (2022) 'Self-Preferencing and Competitive Damages: A Focus on Exploitative Abuses', *The Antitrust Bulletin* 67(2), available at

<https://doi.org/10.1177/0003603X221082757>

Cabral, L., J. Haucap, G. Parker, G. Petropoulos, T. Valletti and M. Van Alstyne (2021) *The EU Digital Markets Act, A Report from a Panel of Economic Experts*, European Commission Joint Research Centre, Luxembourg: Publications Office of the European Union

⁶⁰ Tom Dotan and Miles Kruppa, 'Microsoft, Google, Amazon Look to Generative AI to Lift Cloud Businesses', *The Wall Street Journal*, 27 March 2023, <https://www.wsj.com/articles/microsoft-google-amazon-look-to-generative-ai-to-lift-cloud-businesses-7159a43f>.

⁶¹ Examination of Microsoft's significance for competition across markets, *Bundeskartellamt*, 28 March 2023. The president of the German competition authority stated that "[m]ost recently, the company [Microsoft] has attracted attention with the integration of AI applications. In light of this, there are good reasons to examine whether Microsoft is of paramount significance for competition across markets." See https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2023/28_03_2023_Microsoft.html?nn=3591568.

Carugati, C. (2022) 'How to implement the ban on self-preferencing in the Digital Markets Act', *Policy Contribution* 22/2022, Bruegel

Carugati, C. (2023a) 'The Difficulty of Designating Gatekeepers Under the EU Digital Markets Act', *Bruegel Blog*, 20 February, available at <https://www.bruegel.org/blog-post/difficulty-designating-gatekeepers-under-eu-digital-markets-act>

Carugati, C. (2023b) 'How best to ensure international digital competition cooperation', *Policy Contribution* 02/2023, Bruegel

CMA (2020) *Online platforms and digital advertising*, Market study final report, Competition and Markets Authority, available at https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ALT_T_EXT.pdf

Crémer, J., Y.-A. De Montjoye and H. Schweitzer (2019) *Competition policy for the digital era*, European Commission, Directorate-General for Competition

Digital Competition Expert Panel (2019) *Unlocking digital competition*, HM Treasury, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf

Geng, X., A. Gudiband, H. Liu, E. Wallace, P. Abbeel, S. Levine and D. Song (2023) 'Koala: A Dialogue Model for Academic Research', Berkeley Artificial Intelligence Research, 3 April, available at <https://bair.berkeley.edu/blog/2023/04/03/koala/>

Graef, I., T. Tombal and A. de Streel (2019) 'Limits and Enablers of Data Sharing: An Analytical Framework for EU Competition, Data Protection and Consumer Law', TILEC *Discussion Paper* No. DP 2019-024, Tilburg Law and Economics Center

Gregory, R.W., O. Henfridsson, E. Kaganer and H. Kyriakou (2021) 'The Role of Artificial Intelligence and Data Network Effect for Creating User Value', *Academy of Management Review*, 46(3), available at <https://doi.org/10.5465/amr.2019.0178>

Hoffmann, J., S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford ... L. Sifre (2022) 'Training Compute-Optimal Large Language Models', arXiv:2203.15556, available at <https://doi.org/10.48550/arXiv.2203.15556>

Ji, Z., N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu ... P. Fung (2022) 'Survey of Hallucination in Natural Language Generation', *ACM Computing Surveys* 35(12): 1–38, available at <https://doi.org/10.1145/3571730>

Ouyang, L., J. Wu, X. Jiang, D. Almeida, C.L. Wainwright, P. Mishkin ... R. Lowe (2022) 'Training Language Models to Follow Instructions with Human Feedback', mimeo, Open AI, available at https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf

Paccès, A. and L. Visscher (2011) 'Law and Economics – Methodology', in B. van Klink and S. Taekema (eds) *Law and Method. Interdisciplinary research into Law*, Series Politika, nr 4, Tübingen: Mohr Siebeck

Peitz, M. (2022) 'The Prohibition of Self-Preferencing in the Digital Markets Act', *Issue Paper*, November, Centre on Regulation in Europe

Rae, J.W., S. Borgeaud, T. Cai, K. Millican, J. Hoffmann, F. Song ... G. Irving (2021) 'Scaling Language Models: Methods, Analysis & Insights from Training Gopher', arXiv:2112.11446, available at <https://doi.org/10.48550/arXiv.2112.11446>

Schick, T. and H. Schütze (2021) 'It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners', Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Shapiro, C. (2011) 'Competition and Innovation: Did Arrow Hit the Bull's Eye?' in J. Lerner and S. Stern (eds) *The Rate and Direction of Inventive Activity Revisited*, National Bureau of Economic Research

Vestager, M. (2023) 'Keynote Delivered by EVP Vestager for the Keystone Conference: A Triple Shift for Competition Policy', 2 March, available at https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_23_1342

Wu, Z., M. Sanderson, B.B. Cambazoglu, W.B. Croft and F. Scholer (2020) 'Providing Direct Answers in Search Results: A Study of User Behavior', *CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management*: 1635–1644, available at <https://doi.org/10.1145/3340531.3412017>



© Bruegel 2023. All rights reserved. Short sections, not to exceed two paragraphs, may be quoted in the original language without explicit permission provided that the source is acknowledged. Opinions expressed in this publication are those of the author(s) alone.

Bruegel, Rue de la Charité 33, B-1210 Brussels
(+32) 2 227 4210
info@bruegel.org
www.bruegel.org