Emerging Non-European Monopolies in the Global AI Market

Contact
risto@futureoflife.org

Summary

European companies are not developing general-purpose AI systems and are unlikely to start doing so anytime soon due to their relative competitive disadvantage vis-a-vis American and Chinese players, and due to the amount of money, data, and computational resources it requires to develop these systems. Instead, European companies will likely rely on systems that are developed elsewhere.

The leading general-purpose system, OpenAI's GPT-3 based in San Francisco, has already been integrated in hundreds of downstream applications and last year more than 4.5 billion words per day were generated by customers using it around the world.[1] Unsurprisingly, European companies are among the customers of OpenAI. For example, Netherlands-based MessageBird offers an omni-channel communications platform which is powered by GPT-3, and Finland-based Flowrite turns sequences of words into ready-to-send emails, messages, and posts.[2] Given that the utility of these AI systems largely depends on how many resources are invested and that the biggest players are in the best position to attract foreign capital, these dependencies are likely to grow over time.

---

1   GPT-3 Powers the Next Generation of Apps: https://openai.com/blog/gpt-3-apps/
2   Examples of GPT-3 applications: https://gpt3demo.com/

Main providers

About one year after GPT-3, developed by a San-Francisco company OpenAI, premiered in 2020, a spike in similar model releases followed.[3] These models were developed by private organizations from around the world: Jurassic1-Jumbo, AI21 Labs, Israel; Ernie 3.0 Titan, Baidu, China; Gopher, DeepMind, USA/UK; FLAN & LaMDA, Google, USA; Pan Gu, Huawei, China; Yuan 1.0, Inspur, China; Megatron Turing NLG, Microsoft & NVIDIA, USA; and HyperClova, Naver, South Korea. This suggests that the economic incentives to build such models, and the reputational motivations to announce them, are fairly strong.

So far, all of the cutting-edge large general-purpose AI systems have been developed outside the European Union, as can be seen on this graph from researchers at Anthropic:[4]
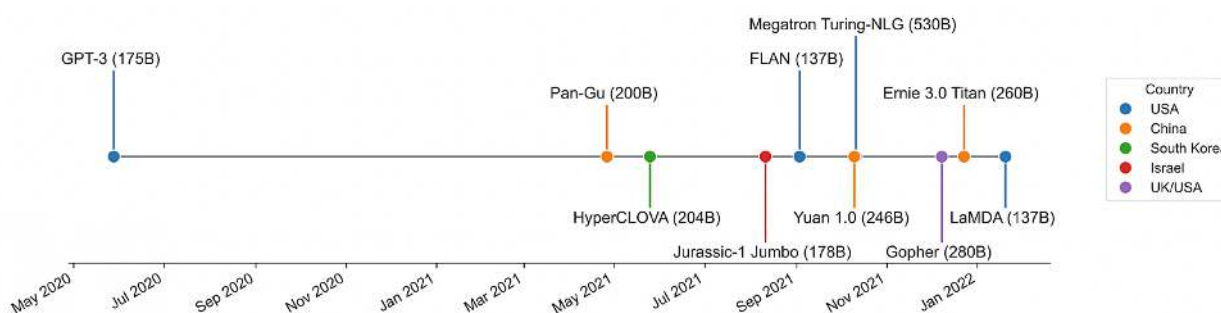
In a similar vein, research by Sevilla and others provides an overview of the world's most powerful AI systems (including narrow, task-specific systems) as displayed on the following graph from Our World in Data. These systems are also developed entirely outside the EU:[5]

Estimated computation used in large AI training runs

Selection of notable AI systems that used a large amount of computation in training. Computation is measured in petaFLOPs, which is 10^{15} floating-point operations.

Source: Sevilla et al. (2022)

Note: The estimates have some uncertainty but are expected to be correct within a factor of ~2.

---

3   Predictability and Surprise in Large Generative Models: https://arxiv.org/abs/2202.07785
4   Predictability and Surprise in Large Generative Models: https://arxiv.org/abs/2202.07785
5   Computation used to train notable AI systems: https://ourworldindata.org/grapher/ai-training-computation?time=2021-06-11..2022-07-01

Market size

According to the latest AI Index report, global private investment in AI totalled $42.57 billion in 2019, $46 billion in 2020, and around $93.5 billion in 2021.[6] The report does not distinguish individual investments in general-purpose AI systems, but other sources do provide some insights.

For example, OpenAI, which focuses exclusively on general-purpose AI systems, formed a multi-year partnership with Microsoft in 2019 valued at $1 billion.[7] DeepMind, another player that focuses almost exclusively on general-purpose AI systems, spent $649M in 2019 alone.[8] Unfortunately, data on Google and Meta's investments in general-purpose AI systems is not publicly available, but – given that they have recently released or advertised models of similar size – their investments are likely comparable to or larger than OpenAI and DeepMind's.[9]

New market entrants are unlikely to raise sufficient capital to compete. Out of the top 100 AI Startups of 2022, only three attempt to build general-purpose AI systems. These startups are Cohere, Hugging Face and AI21 Labs, and are shown in the image below.[10] The only other new market entrant, not yet listed in the top 100, is Anthropic, which was able to raise $580M[11] earlier this year – the largest investment outside the big companies previously mentioned. It is important to note however that Anthropic may more accurately be described as a breakaway company from OpenAI, with the founders of Anthropic all previously working at OpenAI. This may explain how they were able to attract such a large level of funding as a startup. By comparison, in 2022 Cohere raised just $125M,[12] Hugging Face $100M,[13] and AI21 Labs $64M[14].

Pushing the frontier of general-purpose AI models has primarily been restricted to the purview of large corporate entities, because of the need for resources.[15] As a result, the ownership of data and models is often highly centralised, leading to market concentration.[16] To counterbalance this centralisation, there have been grassroots efforts to open source this type of AI development by the likes of Masakhane, EleutherAI, and Hugging Face. However, it is likely that the gap between the private models that industry can train and the ones that are open to the community will remain substantial due to the aforementioned dependence on massive amounts of data and computational resources.

6   Artificial Intelligence Index Report 2022: https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf

7   Microsoft invests $1 billion in OpenAI in new multiyear partnership: https://techcrunch.com/2019/07/22/microsoft-invests-1-billion-in-openai-in-new-multiyear-partnership/

8   DeepMind A.I. unit lost $649 million last year and had a $1.5 billion debt waived by Alphabet: https://www.cnbc.com/2020/12/17/deepmind-lost-649-million-and-alphabet-waived-a-1point5-billion-debt-.html

9   From Meta: https://ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/ and from Google: https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

10   AI 100: The most promising artificial intelligence startups of 2022: https://www.cbinsights.com/research/report/artificial-intelligence-top-startups-2022/

11   Anthropic's quest for better, more explainable AI attracts $580M: https://techcrunch.com/2022/04/29/anthropics-quest-for-better-more-explainable-ai-attracts-580m/

12   Cohere Raises $125M Series B Funding: https://www.finsmes.com/2022/02/cohere-raises-125m-series-b-funding.html

13   Hugging Face raises $100M: https://huggingface.co/blog/series-c

14   OpenAI rival AI21 Labs raises $64M to ramp up its AI-powered language services: https://techcrunch.com/2022/07/12/openai-rival-ai21-labs-raises-64m-to-ramp-up-its-ai-powered-language-services/

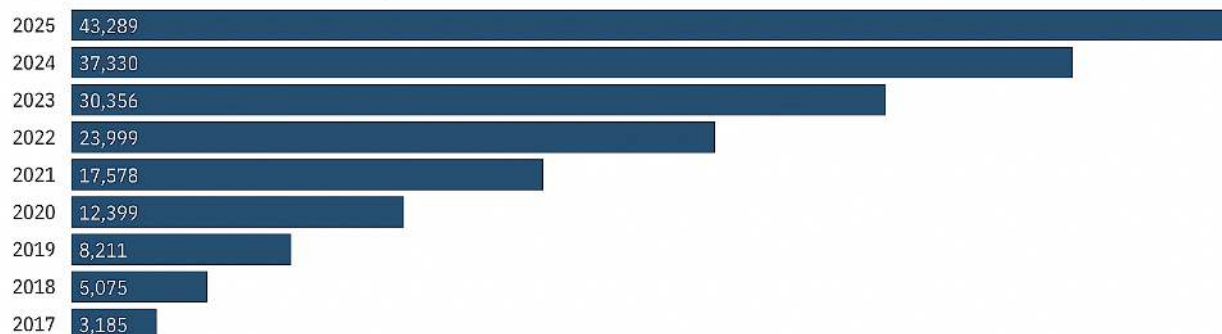15   On the Opportunities and Risks of Foundation Models: https://arxiv.org/pdf/2108.07258.pdf

16   Artificial Intelligence Index Report 2022: https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf
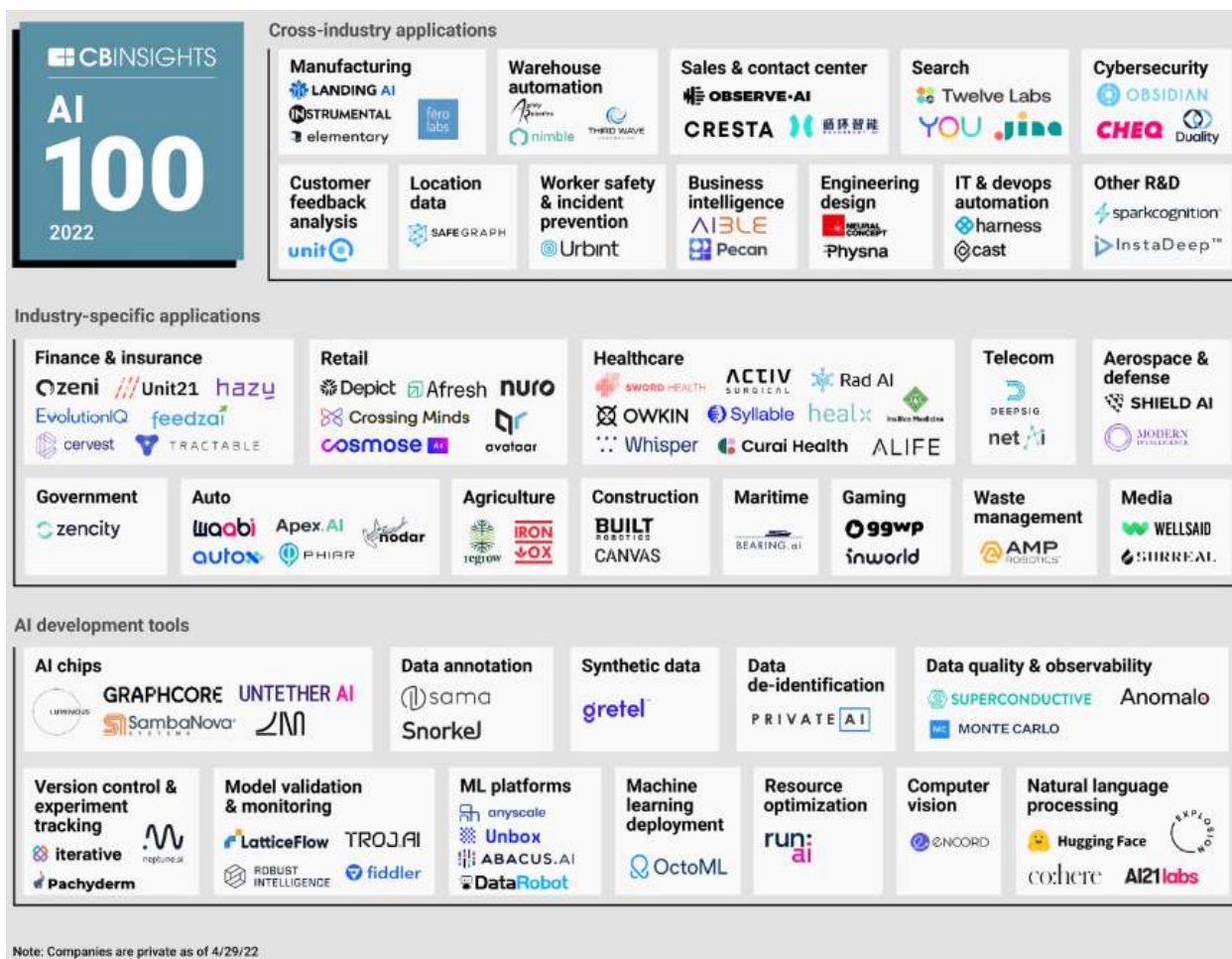
Moreover, the large sums spent on current investment suggest rapid growth of the market for general-purpose AI systems and large language models in particular. According to an estimate by Tech Monitor, systems like GPT-3 could yield 43 billion by 2025:[17]

The NLP market is ready for rapid growth

Systems like GPT-3 could generate $43bn by 2025. Chart shows market size in $M.

Year	Market size ($M)
2025	43,289
2024	37,330
2023	30,356
2022	23,999
2021	17,578
2020	12,399
2019	8,211
2018	5,075
2017	3,185

2020-2025 figures are estimated

Chart: Tech Monitor • Source: Tractica: Natural Language Processing   TECHMONITOR

---

17   OpenAI's GPT-3 influence in the natural language processing field: https://techmonitor.ai/technology/gpt-3-for-businesses-open-ai-microsoft

European examples

We are only aware of two European examples of companies working on general purpose AI systems of any kind. The first is a new project called Large European AI Models (LEAM),[18] which expresses on its website just how difficult it is for small European companies to compete with big American and Chinese developers:

> *"Recent progress made by U.S. and Chinese companies in developing large-scale AI models has the potential to transform the AI market. These larger models will outperform and gradually replace other AI solutions in the market. We fear that European actors will be left behind and the innovative European AI ecosystem will become dependent on American or Chinese solutions."[19]*

The second European example is German startup Aleph Alpha, which raised €28.3 million in 2021 and is trying to develop large language models of the kind that OpenAI has developed, but they are yet to be compared to the state-of-the-art models.[20] Thus, its potential is still to be determined. In any case, small European companies are unlikely to catch up with their global competitors anytime soon.

Costs of development

As we have earlier implied, to understand why general-purpose AI systems are currently mainly provided by a few large companies, it is vital to acknowledge the sheer costs involved. Current state-of-the-art general-purpose AI systems cost more than $9-$35 million to develop.

Lambda Lab estimates that GPT-3 training costs over $4.6 million, research and development costs between $11.4 million and $27.6 million, and the hardware required to run GPT-3 costs between $100K and $150K (without factoring in other costs like electricity, cooling, backup, etc.). On top of all that, running it costs a minimum of $87K per year.[21]

Another set of research indicates that the final training run of AlphaGo Zero in 2017 is estimated to have cost $35M.[22] In addition, Gopher – a frontier NLP model developed in 2021 by DeepMind – cost around $9.2M. PaLM, a new large language model from Google Research, is estimated to have cost between $9 and $17M to train. The training computational cost of developing the next state-of-the-art large AI model will likely be even greater.

Business models

The following is an overview of the main business models of some of the main providers of general-purpose AI systems. Whilst general-purpose AI systems are not limited to language, these types of general-purpose systems do represent the largest group of currently commercially viable products.[23]

All of the following companies offer their systems via an application programming interface (API), which is a software interface to use large pre-trained models for use-case specific applications. Firstly, OpenAI, a company based in San Francisco, has made their general-purpose AI system, GPT-3, available on the market via API for use across applications, charging on a per-word basis. Then

---

18   Europeans' quest to develop AI language models that aren't Chinese or American: https://pro.politico.eu/news/152312

19   LEAM – Large European AI Models initiative: https://leam.ai/

20   German startup Aleph Alpha raises $27M Series A round to build 'Europe's OpenAI': https://techcrunch.com/2021/07/27/german-startup-aleph-alpha-raises-27m-series-a-round-to-build-europes-openai/

21   The GPT-3 economy: https://bdtechtalks.com/2020/09/21/gpt-3-economy-business-model

22   Compute Funds and Pre-trained Models: https://www.governance.ai/post/compute-funds-and-pre-trained-models

23   A Wave Of Billion-Dollar Language AI Startups Is Coming: https://www.forbes.com/sites/robtoews/2022/03/27/a-wave-of-billion-dollar-language-ai-startups-is-coming/

there is Cohere, a fast-growing startup based in Toronto that, like OpenAI, develops cutting-edge natural language processing (NLP) technology and makes it commercially available via API, for use across industries. Based in Israel, AI21 Labs also offers large language models, such as its current state-of-the-art model, Jurassic-1, which powers customers' applications via API. On top of those models, Jurassic-1 also builds and commercializes its own applications.

Another type of business model is a community-based repository for open-source NLP technology offered by Hugging Face. Unlike OpenAI or Cohere, Hugging Face does not develop its own NLP models. Rather, it is a platform that stores, serves and manages the latest and greatest in open-source NLP models, enabling customers to fine-tune these models and deploy them at scale. Customers have to pay a certain fee based on the type of access they want to this platform. It is our understanding that Hugging Face is currently the only company in the world which monetises open-source general-purpose AI systems.[24] Other entities either provide API services, or release open-source models for research purposes.

Meta recently released the large language model OPT-175B open-source.[25] The aim was to get the community to support them in fixing a variety of systematic problems, including their propensity for toxicity. This is likely a one-off attempt to increase public trust, rather than a new business model. Finally, researchers with China's Tsinghua University have recently built and released GLM-130B as open-source, a language model which allegedly outperforms[26] not only OpenAI's original GPT3, but also OPT (Meta's replication of GPT3) and BLOOM (Hugging Face's version).

Future business models

Researchers at the Centre for the Governance of AI predict that an increasing portion of important AI research and development will use large pre-trained models accessible only through APIs.[27] A key reason for this might be that training large models from scratch requires huge resources. Therefore, making these systems available as open-source would make it hard for developers to monetise this large initial investment.

The same researchers provide some evidence that the API business model is starting to gain popularity. For example, they analysed academic papers published at one of the leading natural language processing conferences in 2021 and found that many of the most relevant models are accessible only or primarily through APIs. Similarly, in the market, applications of AI increasingly rely on pre-trained models that are accessed through APIs. Amazon Web Services, Microsoft Azure, Google Cloud, and other cloud providers host pre-trained AI systems for visual recognition and natural language processing on their clouds, and make these available to their customers as a service, via API.

Several other cutting-edge general-purpose systems are not yet commercially available. These include the likes of Google's large language model PaLM and DeepMind's generalist agent Gato.

---

24   Information about Hugging Face's pricing: https://huggingface.co/pricing

25   OPT: Open Pre-trained Transformer Language Models: https://arxiv.org/pdf/2205.01068.pdf

26   Import AI newsletter mentioning the announcement of a Chinese state-of-the-art large language model: https://us13.campaign-archive.com/?u=67bd06787e84d73db24fb0aa5&id=5c97258c8b

27   Compute Funds and Pre-trained Models: https://www.governance.ai/post/compute-funds-and-pre-trained-models