Address The Data Monopoly, Otherwise Tech Giants Control AI

By **Pranav Katti** (September 15, 2023, 3:37 PM EDT)

The growth of artificial intelligence and machine learning over the last several years has been nothing short of extraordinary.

AI is set to revolutionize many fields and industries through its capacity to analyze massive datasets in seconds, make predictions, and learn and adapt without following explicit instructions.

Alphabet Inc. CEO Sundar Pichai believes that AI will be "more profound" to human history than humankind's discovery of electricity or fire.[1]

Pranav Katti

While he has been saying this for years, recent developments like ChatGPT and other sophisticated AI systems have caused the rest of the world to see the unbounded potential of AI.

AI has led to massive technological advances in many fields, with AI and machine learning systems powering personalized recommendation systems, fraud detection tools, medical diagnosis tools, self-driving vehicles, and many more.

Researchers are continuing to push the boundaries and possibilities, addressing even more complex challenges and leveraging AI's potential to solve problems through methods previously deemed unattainable.

We have simultaneously seen over the last several decades the rise of tech companies whose products and services dominate the market and whose use is ubiquitous among the world's population.

Google is the search engine. Amazon.com is the online shopping website. If you walk up to a random person on the street, there is a 50% chance they have an Apple iPhone, regardless of whether they are a member of the middle class or a billionaire.

But there has also been a more subtle monopoly forming. The monopoly on data.

Alphabet, Amazon.com Inc., Apple Inc., Meta Platforms Inc., and Microsoft Corp. are among the most valuable companies in the world.[2]

These companies — whose products are used by billions of people every day — have been able to collect gargantuan amounts of data, ranging from basic information, such as food preferences, locations

and political affiliations, to more intricate information, including the duration one's air-conditioning unit ran, and other minutiae.

With many uses for data and its associated value, it is being called the "new oil."[3] Data is examined through the lens of the four V's: volume, velocity, variety and value. In other words, the quality of the data is an afterthought.

The most important part of a database is the total volume of data.[4] So, the large databases of these tech behemoths are a massive advantage in the development of AI technologies.

Data — and a lot of it — is crucial for teaching AI systems and algorithms.

A large database, among other things, allows for an AI system to improve pattern-matching and relationships in data to make predictions, make complex decisions by learning nuances, adapt to different conditions and transfer what it has learned from one task to another, and even train deep neural networks.

A poorly designed algorithm can find more valuable information and insights in a high volume of data than a superior algorithm can when working with a higher quality but smaller database.[5]

Google's chief scientist, Peter Norvig, even admitted that "[w]e don't have better algorithms than anyone else. We just have more data."[6]

So, it is no surprise that these large tech companies and their databases are leading the journey toward the AI future.

The numbers do not lie. The number of patents related to AI has skyrocketed over the last five years. Not only that, but the owners of the largest number of patents and patent applications are almost exclusively large tech companies.

Where a main purpose of the patent system is to encourage innovation and competition, these numbers are slightly alarming to say the least.

For example, DeepMind, which Google bought in 2014, recently announced that its AI system can predict the structure of every protein in the human body. This is a massive achievement that has the potential to lead to numerous breakthroughs in medical research.

But because inventions created by AI systems are not patentable,[7] the ability, methodology and results of DeepMind's AI to make these predictions is not described in a patent available to the public.

So these potential breakthroughs will only happen if Google allows the medical research community broad access to DeepMind's knowledge. The good news is that Google has decided it will.[8] This time.

This patent and data monopoly has more troubling implications.

In June, the Federal Trade Commission identified their concerns about these tech giants' control of AI. As they noted, a company may engage in "bundling and tying," where they offer a bundle of AI applications and other products together as a single package — bundling — or condition the sale of an AI application on the purchase of separate products — tying.

A company with a large cloud platform may engage in exclusive dealing or discriminatory behavior by steering its own users toward its own AI products instead of competitors' products or give itself preferential cloud computing services that it offers to the company's customers.

A company with vast computational resources for data processing and training — i.e., thousands of graphic processing units[9] — may use mergers and acquisitions to consolidate market power.[10] The list goes on.

These are also no longer theoretical concerns. Two-thirds of all GPUs purchased last year were acquired by either Microsoft, Google, Meta, or Amazon.[11]

This data monopoly is especially interesting because these companies' intellectual property portfolios typically do not encompass the core functionality of their products and services.

Meta's social networking capabilities, Amazon's online retail interface, and Google's search function can all theoretically be recreated and implemented by anybody, and these companies could not do anything about it.

So why are no small players in the tech industry able to usurp these tech giants' status as the overlords of data? The generally accepted answer is what economists call the "network effect."[12]

The network effect occurs where the benefit to a user of a product or service increases as the number of users increases.

Telephones are an example of a network effect, where a telephone is not very valuable when only a few people have them but becomes exponentially more valuable as its user base increases.

In the context of the above-named tech giants, the more data a company accumulates, the better they can make their product or service.

For example, Tesla continuously collects data from its drivers, allowing them to use this data to make their cars even better, and consequently, people start using Tesla's products at an exponential rate. Basically, the network effect causes companies with a high reliance on data to experience positive feedback loops that cause more people to choose their product over others.

But how does this affect AI?

These companies certainly have substantial financial resources to afford the costs of research and development and AI training, they attract the top talent in the field that further asserts their dominance in the market, and they often acquire AI startups at a very early stage, allowing them to acquire these startups' AI systems and patents before these companies become significant competitors.

For example, since 2007, Google has bought at least 30 AI companies working on everything from image recognition to more human-sounding computer voices.[13]

But, at the end of the day, the size of the database is key.

The possession of large databases allows these companies to create more advanced AI models and

continuously improve them, making their products more advanced and therefore more valuable to users.

The network effect arises when these companies collect more data, their AI becomes better, they attract more users, and smaller companies have a harder time competing.

So what can be done? AI and databases have only recently begun to increase in value and be used in the manners described above, so the law is still based on a 20th century understanding of data.

In 2014, IBM Corp. estimated that 90% of the data in the world had been produced in the past two years alone.[14] Currently, the law has tried to handle these issues with antitrust and consumer law, but both are poorly equipped to fully address data monopolization because they only address the results of the problem and not the origin: the monopoly itself.

Ultimately, whether the change comes from the legislature, the courts or tech companies, the way we treat data needs to change.

When a dataset becomes so large and ubiquitous that it encompasses a wide range of fields of human innovation, then the controller of that data becomes a very powerful force that is capable of controlling the direction of humanity's future.

One possible solution is imitating the rules of the patent system. We recognize that inventions are useful, and while we want to encourage innovation, we want to allow access and disclosure of these inventions to the public.

So, we allow exclusive use of patented inventions for a certain term, after which the invention is free to use by the public. Data could be similarly controlled by a legal mechanism that enforces a time limit for exclusive use of the data.

Collected data would be given a limited term, could be used exclusively during that term, and once expired, the controller of the data would be obligated to make the data available for public use.

Essentially, when a time-stamped block of data reaches a certain age, it is automatically publicized for open access.[15] Currently, there is little reason for a company to share its data. A system like the above would create such an incentive and subject data to the same regulating mechanism for intellectual property: its term.

A second possibility is offering tax incentives to companies for data sharing. Offering incentives or tax breaks to companies that share certain types of data with the research community or smaller startups, which would encourage data collaboration, would be a simple way to address this data monopoly.

A third possibility, offered by Oxford internet governance professor Viktor Mayer-Schönberger, has called for a mandate for technology companies to make more data available to the public as their market share increases.

According to Sunny Seon Kang, senior privacy counsel and head of policy at Inpher:

> [The mandate] would kick in once a company's market share reaches an initial threshold — say, 10 percent. It would then have to share a randomly chosen portion of its feedback data with every other player in the market that requests it. How much data it must make available would depend

on the market share captured by the company. The closer a company is to domination, the more data it would have to share with its competitors.[16]

On a positive note, there has been a notable change in Europe in the last several years. There is a growing acknowledgment that the control of databases can be used as a barrier to entry.

Last year, the European Commission gave Google a record-breaking €2.42 billion fine for abusing its dominance as a search engine by giving preference to its own services offered to users in the online shopping market.[17]

The General Data Protection Regulation also provides citizens various rights in enforcing their data rights, like the right of access to their personal data, the right to be forgotten, a requirement that consent to collect personal data be freely given, a right to notification of any data breach, and a right not to be subject to decisions solely based on automated processing.

Significantly, the GDPR introduces a "data portability right," which gives a person the right to be able to entirely transfer their personal data from one provider to another.[18] Data portability protect users from having their data stored in systems that are incompatible with one another, thus subjecting these users to vendor lock-in and the practical inability to move to another product or service.

This is good for Europe, but seeing as how half of AI inventions and products come from the U.S.[19] something needs to change, whether we follow Europe's lead or come up with our own solution.

Unless this change happens, it is likely that we will experience a severe monopoly on AI systems and patents by the largest players in the tech industry.

Pranav Katti is an associate at Barclay Damon LLP.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] Prarthana Prakash, Alphabet CEO Sundar Pichai says that A.I. could be 'more profound' than both fire and electricity—but he's been saying the same thing for years (2023), available at https://fortune.com/2023/04/17/sundar-pichai-a-i-more-profound-than-fire-electricity/.

[2] Section on Largest Companies by Market Cap, https://companiesmarketcap.com/.

[3] The New Oil: Data Is the World's Most Valuable Resource, AUSTL. (2017), available at https://www.theaustralian.com.au/news/inquirer/the-new-oil-data-is-the-worlds-most-valuable-resource/news-story/f386217a9c63ac5ee6el473413e90bda. 108. JAMES MANYIKA ET AL., supra note 93, at 2.

[4] See Andrea De Mauro et al., A Formal Definition of Big Data Based on Its Essential Features, 65 LIB. REV. 122, 130-31 (2016). ("Analyst Douglas Laney first defined the three "v"s of volume, velocity and variety. Andrea De Mauro, Marco Greco, and Michele Grimaldi added the "value" component.").

[5] Maurice E. Stucke & Allen P. Grunes, Data-opolies 8, 11 (Univ. of Tenn. Legal Studies Research, Paper

No. 316, note 37 at 23, 2017).

[6] Matt Asay, Tim O'Reilly: "Whole Web" Is the OS of the Future, CNET (2010), available at www.cnet.com/news/tim-oreilly-whole-web-is-the-os-of-the-future/ (quoting Peter Norvig, chief scientist at Google).

[7] See Thaler v. Vidal, 43 F.4th 1207 (Fed. Cir. 2022).

[8] Bhaskar Chakravorti, Big Tech's Stranglehold on Artificial Intelligence Must Be Regulated (2021), available at https://foreignpolicy.com/2021/08/11/artificial-intelligence-big-tech-regulation-monopoly-antitrust-google-apple-amazon-facebook/.

[9] Section on GPUs, https://www.intel.com/content/www/us/en/products/docs/processors/what-is-a-gpu.html ("GPUs were originally designed to accelerate the rendering of 3D graphics. Over time, they became more flexible and programmable, enhancing their capabilities. This allowed graphics programmers to create more interesting visual effects and realistic scenes with advanced lighting and shadowing techniques. Other developers also began to tap the power of GPUs to dramatically accelerate additional workloads in high performance computing (HPC), deep learning, and more. Because GPUs incorporate an extraordinary amount of computational capability, they can deliver incredible acceleration in workloads that take advantage of the highly parallel nature of GPUs, such as image recognition. Many of today's deep learning technologies rely on GPUs").

[10] Staff in the Bureau of Competition & Office of Technology, Generative AI Raises Competition Concerns (2023), available at https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns.

[11] Arthur Sants, How cloud computing became a global monopoly (2023), available at https://www.investorschronicle.co.uk/news/2023/05/09/how-cloud-computing-became-a-global-monopoly/.

[12] Mark A. Lemley & David McGowan, Legal Implications of Network Economic Effects, 86 CAL. L. REV. 479, 481 (1998).

[13] Bhaskar Chakravorti, Big Tech's Stranglehold on Artificial Intelligence Must Be Regulated (2021).

[14] Todd Vare & Michael Mattioli, Big Business, Big Government and Big Legal Questions, MANAGING INTELL. PROP., Oct. 2014, at 46, available at http://www.managingip.com/Article/3382483/Bigbusiness-big-government-and-big-legal-questions.html.

[15] McIntosh, Daniel (2019) "We Need to Talk About Data: How Digital Monopolies Arise and Why They Have Power and Influence," Journal of Technology Law & Policy: Vol. 23: Iss. 2, Article 2. Available at https://scholarship.law.ufl.edu/jtlp/vol23/iss2/2.

[16] Sonny Seon Kang, Don't Blame Privacy for Big Tech's Monopoly on Information (2020), available at https://www.justsecurity.org/72439/dont-blame-privacy-for-big-techs-monopoly-on-information/.

[17] Ricardo Cardoso, Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service (2017), available at https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1784.

[18] McIntosh, Daniel (2019) "We Need to Talk About Data: How Digital Monopolies Arise and Why They Have Power and Influence."

[19] Imane El Atilllah, Artificial intelligence race: These are the countries developing top AI tools (2023), available at https://www.euronews.com/next/2023/06/15/artificial-intelligence-race-these-are-the-countries-developing-top-ai-tools#:~:text=More%20than%20half%20of%20the,of%20the%20top%20AI%20tools.