30 September 2018

# MICROSOFT RESPONSE

# EC CONSULTATION ON COMPETITION POLICY IN THE ERA OF DIGITISATION

Microsoft welcomes the efforts of the European Union Commissioner of Competition, Margrethe Vestager, to convene a panel of external expert advisors and a conference to provide input on how competition policy can best serve European consumers in our fast-changing world.

Over the past 20 years, digital technology has transformed the way we live and work. Digital technology powered by the cloud has made us smarter and helped us optimize our time, be more productive and communicate with one another more effectively. And this is just the beginning. But, as we have seen over the past 20 years, as digital advances bring us daily benefits they also raise a host of complex questions and broad concerns about how technology will impact society. And this seems certain to continue as artificial intelligence (AI) evolves and as commerce is increasingly facilitated by online platforms. Governments, and particularly competition regulators, are right to ask whether they are taking the right steps to ensure healthy and vibrant competition and innovation in the increasingly digital economy. Microsoft appreciates the opportunity to contribute our thoughts and insights on the questions being posed and considered by the Commissioner.

## *Microsoft's Approach to Artificial Intelligence*[1]

Artificial intelligence – and the data that drives it – is at the forefront of questions, including those posed by competition regulators, about the digital economy. At Microsoft, we think of AI as a set of technologies that enable computers to assist in decision-making to solve problems in ways that are similar to people. With these capabilities, how computers understand and interact with the world is beginning to feel far more natural and responsive than in the past, when computers could only follow pre-programmed routines.

Researchers have been working on AI for decades. Progress has accelerated over the past few years thanks in large part to three developments: the increased availability of data; growing cloud computing power; and more powerful algorithms developed by AI researchers.

As our lives have become increasingly digitized and sensors have become cheap and ubiquitous, more data than ever before is available for computers to learn from. Analyzing all this data requires massive computing power, which is broadly available to all thanks to the efficiencies of

[1] *See also* Smith, Brad and Harry Shum, The Future Computed: Artificial Intelligence and its Role in Society (17 Jan 2018) (detailing Microsoft's approach to artificial intelligence) (available at https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf). Additional information on Microsoft's approach to artificial intelligence is available at https://www.microsoft.com/en-us/ai?activetab=pivot1%3aprimaryr5). Microsoft makes available publicly information regarding its AI research at https://www.microsoft.com/en-us/research/research-area/artificial-intelligence/.

cloud computing. Today, organizations of any type can tap into the power of the cloud to develop and run their AI systems.

Researchers at Microsoft, other technology firms, universities and governments have drawn upon this combination of the availability of this data, and with it ready access to powerful computing and breakthroughs in AI techniques — such as "deep learning" using so-called "deep neural nets"— to enable computers to mimic how people learn.[2]

Today, Microsoft is building AI systems that are designed to amplify natural human ingenuity. Broadly speaking, the kind of "computational intelligence" that computers can provide will have a significant impact in almost any field where intelligence itself has a role to play. AI systems will help people tackle big problems, the most complex and pressing issues that humanity faces: from reducing poverty and improving education, to delivering healthcare and eradicating diseases, addressing sustainability challenges such as growing enough food to feed our fast-growing global population through to advancing inclusion in our society. For example, researchers at Microsoft have teamed up with oncologists to develop an AI system to help treat cancer more effectively.[3] And they are working with university scientists and engineers to detect pathogens in the environment so public health officials can protect people from transmission before an outbreak begins.[4] But Microsoft cannot deliver on the promise of AI unless we make it broadly available to all. So, we continue to encourage our researchers to publish their results broadly so that AI researchers around the world — at universities, at other companies and in government settings — can build on these advances. For our customers, we're building AI capabilities into our most popular products, such as Windows and Office. And we are also making them available to developers so that they can build their own AI-powered products. The Microsoft AI Platform offers services, tools and infrastructure making AI development easier for developers and organizations of any size.

In many ways, AI is still maturing as a technology. Most of the progress to date has been in teaching computers to perform narrow tasks — play a game, recognize an image, predict traffic. We have a long way to go to imbue computers with "general" intelligence. Today's AI cannot yet begin to compete with a child's ability to understand and interact with the world using senses such as touch, sight and smell. And AI systems have only the most rudimentary ability to understand human expression, tone, emotion and the subtleties of human interaction. In other words, AI today is strong on "IQ" but weak on "EQ."

It's safe to say that most current standards, laws and regulations were not written specifically to account for AI. But, while existing rules may not have been crafted with AI in mind, this does not mean that AI-based products and services are unregulated. Current laws that, for example, protect the privacy and security of personal information, that govern the flow of data and how it is used, that promote fairness in the use of consumer information, or that govern decisions on credit or employment apply broadly to digital products and services or their use in decision-

---

[2] For more information on deep learning *see, e.g.,* Andrej Karpathy, "Software 2.0," (11 Nov. 2017) (available at https://medium.com/@karpathy/software1-0-a64152b37c35).

[3] *See* https://news.microsoft.com/stories/computingcancer/.

[4] *See* https://www.microsoft.com/en-us/research/project/project-premonition/.

making, whether they explicitly mention AI capabilities or not. AI-based services are not exempt from the requirements that will take effect with GDPR, for example, or from HIPAA regulations that protect the privacy of healthcare data in the United States, or existing regulations on automobile safety.

As the role of AI continues to grow, it will be natural for governments not only to monitor its impact, but to address new questions and update laws. One goal should be to ensure that governments work with businesses and other stakeholders to strike the balance that is needed to maximize the potential of AI to improve people's lives and address new challenges as they arise. As we look to the future, it's important that we maintain an open and questioning mind while we seek to take advantage of the opportunities and address the challenges that this new technology creates.

***Addressing the Role of Data***

The development of more effective AI services requires the use of data – often as much relevant data as possible. As such, competition regulators here and across the world are focused on the its collection and use. As vast amounts of data are generated through the use of smart devices, applications and cloud-based services, some competition regulators have begun to explore the possibility of data "bottlenecks" or whether data and information could become concentrated in a relatively small number of companies.

Competition regulators should monitor whether it is possible for access or control over unique datasets (in other words, data for which there is no substitute) to become a barrier to competition. At the outset, however, it is worth noting that many types of data are replicable; consumers can choose to use and provide the same data to more than one online service. Such "multi-homing" means similar data may be easily available. And companies can and do generate proprietary data; companies, for example, collect "observed" or "inferred" data in the process of providing an online service. That data may be used to improve the online service by, for instance, making the user experience more intuitive and easier or by predicting what offers and products the user may be most interested in. But that data is not necessarily "unique;" a competing online service can generate the same type of data vis a vis its usage. While some may view the technical ability to distill and process large amounts of data as a competitive problem, that ability is the result and driver of innovation and investment that competition regulators should incentivize.

If unique datasets are present, competition regulators can address concerns about access to those datasets under the framework of existing competition law. Indeed, competition law principles established over decades have been applied successfully to a wide variety of changing fact patterns, and they can be successful applied to data as well. From a competition perspective, data are simply a class of assets that vary in their competitive significance. They can be a product, an input for some other product, or commercially irrelevant. The challenge for competition regulators is to separate cases requiring closer scrutiny from the bulk of cases where data control and usage is economically beneficial, drives innovation and is competitively benign.

Historically, the European Union has set a high bar for compelling access to proprietary assets;[5] data should be no different. Collecting and using proprietary data – even in large amounts – to seek competitive advantage is economically efficient behavior that drives innovation. This is true, even when the competition is *for* the market rather than *within* the market. Competition for the market is common in online consumer services, which are often two-sided markets characterized by network effects. Economic theory and practical experience demonstrate that such competition is intense, with competitors striving to develop leap-frog technology that becomes the solution most people choose. And the promise of the rewards that come with winning the market creates an incentive for firms to take outsized risks to invest in the development of these solutions.[6] Mandating access to "nice to have," or even necessary, data to compete within a market reduces that incentive and it relieves the burden on competitors to innovate on their own, creating new, better products and solutions. And any obligation to share data – particularly to the extent that such data involves personally identifiable information – may implicate privacy concerns and impinge on the rights and freedoms of natural persons. Competition regulators should be skeptical of calls to regulate access to data and take extreme care to avoid false positives. And most importantly, they should resist the temptation to engineer market outcomes that protect parochial interests and stray from protecting the competitive process.

Theoretically, a competitor *might* obtain control over data essential to competition in a relevant market. But, given the role of data in the development of AI services, it is rarely the case that a competitor would have exclusive control over *all* sources of necessary data and any substitutes.

While machine learning techniques generally rely upon large data sets for training algorithms, the utility of a machine learning model depends on its ability to deliver predictive insights that are not obvious from any single source of data. As such, it is the multi-dimensionality of the data inputs and the signals that they provide that matters; simply having more of the same data is not enough. Valuable insights are gained from assembling and interrogating heterogeneous sets of data across multiple categories with sophisticated analytics to deliver unique insights and better explanatory power. It is rare that a single entity will control all relevant datasets.

---

[5] *See* "Guidance on the Commission's Enforcement Priorities in Applying Article 82 of the EC Treaty to Abusive Exclusionary Conduct by Dominant Undertakings," para. 83 (2009/C 45/02) ("Rather, an input is indispensable where there is no actual or potential substitute on which competitors in the downstream market could rely so as to counter – at least in the long-term – the negative consequences of the refusal. In this regard, the Commission will normally assess whether competitors could effectively duplicate the input produced by the dominant undertaking in the foreseeable future."). *See also* Case T-167/08 *Microsoft v. Commission* ECLI:EU:T:2012:323 (finding a refusal to license IP abusive if it is imposed by a dominant company, it eliminates effective competition in a secondary market, it prevents the emergence of a new product or limits technical development, and it is not objectively justified); Case C-418/01 *IMS Health v. NDC Health* [2004] ECR I-5039, para. 38.

[6] *See* Nigro, Bernard A., "Big Data and Competition for the Market," Remarks as Prepared for Delivery at The Capitol Forum and CQ: Fourth Annual Technology, Media, & Telecom Competition Conference (13 Dec 2017) (available at https://www.justice.gov/opa/speech/file/1017701/download). *See also* Competition & Markets Authority, "Just Eat and Hungry house: A report on the anticipated acquisition by JUST EAT plc of Hungryhouse Holdings Ltd." at App. E (16 Nov. 2017) (providing an example of how network effects operate in a two-sided market for online food platform) (available at https://assets.publishing.service.gov.uk/media/5a0d64fee5274a0ee5a1f22a/appendices-glossary-final-report-justeat-hungryhouse.pdf).

Moreover, not all data is commercially useful. The better the structure, velocity, variety, density, and volume of relevant data, the better the opportunities to identify patterns and make useful predictions. However, whether a type of data is relevant – let alone indispensable – can only be determined by considerable statistical study. It is often not obvious in advance whether the addition of a given data set will add more predictive ability. In many cases, a company's own proprietary data is likely to be the most valuable dataset in training a machine learning model. Whether or not third-party data sets – acquired via merger or license agreement – offer additional incremental predictive value is unclear; it depends on how correlated the dataset is with consumer behavior related to that business. Generally, no given data set is itself so important that it could tip the market given the variety of data that is required and the alternatives available to achieve that variety.

To evaluate concerns regarding access to datasets – whether in the context of a potential merger or abuse of dominance – Microsoft proposes this simple framework, largely taken from the 2017 article "*Is Big Data a Big Deal?*":[7]

- **Do the parties own or control the relevant data?** The potential competitive significance of a given dataset will depend to a significant degree on who is the controller (within the meaning of the data protection laws)[8] for that data. Data is useful only if it can be accessed and used. A controller may use its own data and, typically with user consent, may use volunteered or observed data[9] for specified purposes, which may include the development and improvement of its AI services. In contrast, a processor may access data but only as instructed by the customer (controller) and not for its own use. Accordingly, data for which a party merely acts as a processor is not competitively significant in markets where that data processor competes. Similarly, there are no significant competition concerns in the markets for software and services that help owners of datasets collect, store, organize, and analyze their own data locally or "on-premises;" such tools do not convey any special "data asset" to the vendor of those tools.
- **Is the relevant data commercially available as a product or as an input for products of downstream competitors?** When data is sold as a product, competition concerns should only arise if the company controlling the data can realistically use it to foreclose competition in a downstream market, *i.e.,* a market for a product that is built using that data set.[10] When data is not sold as a product or not otherwise available to competitors, it is generally unlikely to represent a unique critical input in any competitive sense. In fact,

---

[7] Greg Sivinski, Alex Okuliar & Lars Kjolbye (2017) "Is big data a big deal? A competition law approach to big data, European Competition Journal," 13:2-3, 199-227, DOI: 10.1080/17441056.2017.1362866 (available at https://www.tandfonline.com/doi/full/10.1080/17441056.2017.1362866).

[8] *See, e.g.,* General Data Protection Regulation (2016/679) (Article 4, paras. 7-8, defining controller and processor with respect to personal data).

[9] "Volunteered data" is shared intentionally, such as personal information provided by a consumer to his or her online banking provider. "Observed data" is obtained by tracking user activity in software or online, such as user clicks on online ads.

[10] *See, e.g.,* Competitive Impact Statement, *U.S. v. Google, Inc., and ITA Software, Inc.*, No. 1:11-cv-688 (D.DD.C. 8 April 2011) (available at https://www.justice.gov/atr/case-document/file/497671/download) (explaining that the Final Judgment required Google to make ITA available to competitors because it was essential to competition in the downstream comparative flight search market).

if the data controller has downstream competitors (those that develop a product without using the data), that strongly suggests that those competitors have access to comparable internal or external data sets.

- **Is the relevant data proprietary and captive to the owner or controller's own products or services?** Some AI services are built using large amounts of data. Using data to seek a competitive advantage is economically efficient behavior that drives innovation; it is difficult to determine when such behavior becomes illegal exclusionary conduct. If data is proprietary – not shared with others or only shared under certain conditions that protect its proprietary nature – then it is highly unlikely to be necessary to compete in the downstream market.
- **Do reasonably available substitutes for the relevant data exist or is it unique?** In most cases, the data is useful but not essential to compete or there are reasonable substitutes. Many data already have or could have full or partial substitutes that can be combined or can be collected by starting new lines of business. Only in rare cases is a given data set necessary for competition in each market and without reasonable substitutes.

In short, before taking any action based on a company's use or acquisition of sets of big data, competition regulators should proceed with great caution to understand in depth what data is relevant in each case, how that data is used, and whether any substantial foreclosure is indeed possible.

***The Challenge of Transparency***

AI systems are increasingly being used to make consequential decisions that affect consumers and businesses—decisions that humans would have made in the past. For example, an AI system may largely replace a travel agent in determining whether to promote one airline's flight over another, or it may constantly survey online prices for a particular good and make adjustments to the price at which it is offered to stay competitive. The automated nature of these considerations and their potential impact on competition leads some competition regulators to ask how AI systems make decisions.

As a first step, the new European General Data Protection Regulation (GDPR) strengthens transparency requirements and accordingly firms processing personal data need to make AI more "explainable" or "transparent." We may see additional legal requirements over time. Microsoft is working with industry, policymakers, and governments to develop a workable approach that will satisfy legal requirements, our customers, and the public at large.

Simply publishing the algorithms underlying AI systems will rarely provide meaningful transparency. With the latest (and often most promising) AI techniques, such as deep neural networks, there typically is no algorithmic output that would help people understand the subtle patterns that systems find. And disclosing nothing more than granular code underlying a complex AI system will not ensure that that system is operating fairly or predictably. Such an understanding is only possible if the AI system is considered in its broader context, and there is an attempt made to provide information about the overall system composition rather than the underlying code or data sets.

An approach that is most likely to engender trust with users and those affected by these systems is to provide explanations that include contextual information about how an AI system works and interacts with data. Such information will make it easier to identify and raise awareness of potential bias, errors and unintended outcomes. Microsoft is working with the Partnership on AI and other organizations to develop best practices for enabling meaningful transparency of AI systems without undermining incentives to innovate. This includes the practices described above and a variety of other methods, such as an approach to determine if it's possible to use an algorithm or model that is easier to understand in place of one that is more complex and difficult to explain. This is an area that will require further research to understand how machine learning models work and to develop new techniques that provide more meaningful transparency.

***Spurring Innovation***

Today's digital technology enables faster and more profound progress in nearly every field of human endeavor, and it is essential to enabling the digital transformation that is at the heart of worldwide economic development. Yet competition regulators are appropriately asking whether competition and innovation will continue to flourish in the future, and what steps are necessary to ensure that they do. Mergers in the technology industry and their potential impact on innovation have become a particular point of scrutiny.

Competition regulators can and do routinely evaluate the impact of a proposed merger on the incentives for innovation: does a proposed merger eliminate the "racing dynamic" between companies engaged in direct product competition and pursuing similar research and development (R&D) to "get there first?" As the Competition Directorate-General of the European Commission observed in its 2016 Competition Policy Brief "[t]he EU's legal framework for merger control explicitly addresses a merger's effects on innovation – either positive or negative . . . ."[11]

Innovation effects are typically most relevant and concerning in horizontal mergers in concentrated product markets with well-established research pipelines where innovation is path-dependent. In these instances, innovation tends to be directed, horizontal, and predictable rather than undirected, vertical, and unpredictable. And evidence of durable R&D pipelines and physical and intellectual property assets and capabilities that may be barriers to entry can be easily identified. Data on patents may be used to infer the structure of innovation and data on pipelines and the wealth of those pipelines may allow for structural inferences.

Innovation in the technology industry can be disruptive or leapfrog innovation is more common, and barriers to innovation are often low. Moreover, innovation is incremental and often arises organically from a single engineer or small group of software developers working within a defined product team such as Office 365. Well-established research pipelines, much less a formal R&D structure, may not even exist or may be focused upon unrelated technologies. As a result,

[11] Competition Directorate-General, "EU Merger Control and Innovation," Competition Policy Brief at 7 (April 2016) (available at http://ec.europa.eu/competition/publications/cpb/2016/2016_001_en.pdf).

identifying or (if the burden is placed on the merging entities) disproving potential innovation effects is challenging and fraught with peril. The risk of false positives is high.

Indeed, the history of mergers in the technology industry is littered with mistakes, proving that it is exceedingly difficult – even for business executives and engineers steeped in the industry – to predict with accuracy what combinations will succeed and prove to be consequential. For example, in 2005, eBay purchased Skype for $2.6 billion, with the aim of enabling VoIP and video calls between its online buyers. But online buyers were not interested and after four years, eBay sold Skype at a loss to private investors for $1.9 billion. And, in 2014, Microsoft purchased Nokia's devices business for $7.6 billion to accelerate its Windows Phone only to record a multi-billion loss less than two-years later and dramatically scale back its mobile phone business. In contrast, in 2005, Google purchased Android for an undisclosed price (estimated at $50 million). Few could have predicted that that it would become one of the largest operating systems in the world.

Imposing unique rules, constraints, or presumptions on mergers – whether horizontal or vertical – in digital markets risks over intervention that chills innovation. Competition regulators should hesitate to take such a risk when there is little evidence that the rate of innovation is declining or will decline in the technology industry.

The fact that large technology companies routinely acquire small, start-up technology firms does not suggest otherwise. To the contrary, these acquisitions are a key ingredient to spurring innovation in the industry.[12] Indeed, creating a vibrant and healthy pipeline of start-ups depends in part on the potential for acquisition by relatively larger technology firms. Start-ups are dependent on venture capital for their existence. The potential for acquisition – which typically generates a profitable return – incents venture capitalists to invest in start-ups. Additionally, starts-ups often lack the resources, skills, and capabilities to commercialize their innovative products and services at scale and meet the wide-ranging demands of various customer groups (*e.g.*, enterprises, governments, educational institutions, healthcare entities). Acquisitions by a relatively larger technology firms – with large-scale and sophisticated development, operations, marketing, and sales teams – can help ensure that these innovative products and services are brought to market.

To spur innovation, competition regulators can and should lend their expertise to advocate for and help design broad policies and programs that support digital transformation across a wide range of industries and businesses of all sizes. For example, governments can help accelerate AI advances by promoting common approaches to making data broadly available for machine learning. A large amount of useful data resides in public datasets — data that belongs to the public itself. Governments can help add to the supply of available data by ensuring that public data is usable by AI developers on a non-exclusive basis. In addition to addressing issues relating

[12] *See, e.g.,* D. Daniel Sokol, "Vertical Mergers and Entrepreneurial Exit" (16 July 2018) (concluding vertical merger policy that would unduly restrict large tech firms from undertaking acquisitions would hurt incentives for innovation in the economy by chilling business formation in startups.) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3217095).

to data, governments have an important role to play in promoting responsible and effective uses of AI itself. And governments have an important role to play in funding core research to further advance AI development and support multidisciplinary research that focuses on studying and fostering solutions to the socioeconomic issues that may arise as AI technologies are deployed. This multidisciplinary research will also be valuable for the design of future AI laws and regulations.

***Conclusion***

The existing framework of competition law and its analytical tools are appropriate, mature, and robust enough for addressing the challenges to competition presented by digital technology. Indeed, both have been developed over decades and have historically proven to be capable of and relevant to addressing a variety of economic and industrial changes. As with all change, the challenge is to understand the technology and competitive dynamics at play to accurately separate cases requiring closer scrutiny from the bulk of cases where the behavior is economically beneficial, drives innovation and is competitively benign. Microsoft remains available to support this consultation and engage in further discussions to explain the technology as well as market trends underlying the continuing digitisation of the economy.