RETHINKING DATA AS AN ESSENTIAL FACILITY IN AI-DRIVEN MARKETS: A NARROW APPROACH FOR INDIAN COMPETITION LAW

Aadya Mishra, Maharashtra National Law University, Mumbai

INTRODUCTION

The past decade has seen India experience an unprecedented rate of digitalization in its transformation journey,[1] with emerging technologies like artificial intelligence, a platform-based economy, and data-driven governance significantly impacting both the market and governance systems in the country.[2] The legislative response has been to try and keep up with these changes, as evident in the Digital Personal Data Protection Act, 2023;[3] however, there has been a noticeable gap in the ability of the institutions tasked with implementation compared to the design of regulatory responses.[4] This is not just a question of implementation; rather, there are fundamental questions about whether India's regulatory system is fit for its role in governing these emerging technologies.

The majority of existing research has concentrated on issues that are critical to regulation, such as privacy, surveillance, and accountability. This has largely been done using concepts that have already been developed elsewhere, such as those that have been developed within regions such as the EU, especially under the General Data Protection Regulation.[5][6] However, there appears to be little attention given to the regulation environment within India. This includes an examination of how the regulation environment interprets, applies, and enforces technology regulations, as they struggle to cope with their own capabilities. This paper argues that the main challenge within India is not that there are no regulations, but that there is a lack of regulatory capabilities that are required to make these regulations effective governance.

[1] OECD, DIGITAL ECONOMY OUTLOOK 2023 (2023).
[2] NITI Aayog, NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE (2018).
[3] The Digital Personal Data Protection Act, 2023 (India).
[4] Graham Greenleaf, Global Data Privacy Laws 2023, 179 PRIVACY LAWS & BUS. INT'L REP. (2023).
[5] General Data Protection Regulation (EU) 2016/679, 2016 O.J. (L 119) 1 (EU).
[6] ORLA LYNSKEY, THE FOUNDATIONS OF EU DATA PROTECTION LAW (2015).

At the same time, global regulatory models offer instructive contrasts. This shows us that there are different ways to adapt to what can be done. The central approach taken by the GDPR, as well as the flexible approach taken by Singapore, are useful examples. However, it is not possible to use these examples to discuss India because the roles of the regulatory agencies have been broken, and there is a lack of technical knowledge.

So, this paper wants to look into how policy design and regulatory capacity are connected in India's technology governance framework. It looks into how institutional limits affect regulatory outcomes and shows that we need a more unified approach to deal with the problems that come with rapid technological progress.

RETHINKING MARKET POWER IN DATA-DRIVEN AND AI MARKETS

In order to fully understand this phenomenon of data and its impact on competition, we are required to look away from those traditional concepts of how a firm can achieve dominance in a market. In the digital world, dominance in data not only helps to augment but also build market dominance. However, this phenomenon is not applicable to all sectors but occurs in a very specific manner that differentiates data dominance from traditional forms of economic dominance.

The first way in which data dominance occurs is in the form of data-driven network effects.[7] While traditional network effects depend directly on the number of users to create value, data-driven network effects are more indirect but create similar levels of value. When users interact with a service or product, they create huge levels of data in the form of think search queries, click patterns, and preferences.

This data, in turn, can be used to further improve and enrich service quality even more. Thus, improved service quality results in an increase in the number of users, which in turn results in an increase in data. This creates a self-reinforcing cycle in which initial advantages are further enhanced. [8]Another aspect to be considered is the initial advantage in terms of behavioral data.[9]

[7] GEOFFREY G. PARKER, MARSHALL W. VAN ALSTYNE & SANGEET PAUL CHOUDARY, PLATFORM REVOLUTION (2016).
[8] Hal R. Varian, Artificial Intelligence, Economics, and Industrial Organization (2019); OECD, Algorithms and Collusion: Competition Policy in the Digital Age (2017).
[9] SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM (2019).

Not all data can provide equal benefits in terms of increasing and enhancing competitiveness.[10] What are most important are those pieces of data that provide real-time data about user behavior over a long period of time. This data is extremely context dependent. This data is extremely dynamic. This data is extremely proprietary. This data, unlike static data, cannot be easily replicated by any newcomer. This gives an initial qualitative advantage to any organization that's already in possession of such data. This initial qualitative advantage is extremely difficult to beat through conventional means of competition.

The rising trend of artificial intelligence is definitely making waves. Artificial intelligence systems, particularly the ones that utilize machine learning technology, are largely dependent on the availability of larger datasets in order to improve and enhance these systems. The relationship between data and performance in these systems is complex and not immediately obvious. What is known is that the more data that is available, the better the performance of these systems.[11] This has created a rather fascinating feedback loop in the artificial intelligence market.[12] The companies that are able to utilize more data are able to create more efficient and more accurate systems. These systems are able to attract more and more users, thereby generating even more data. This eventually leads to a performance gap that is very difficult for competitors to overcome, even if they possess similar expertise in the field.

However, we should not overstate the importance of data in all cases. Indeed, not all data sets will offer such advantages.[13] In many cases, data sets may be readily available, easily replicable, or even substitutable. Public data sets, common knowledge, and transitory data streams are not likely to offer a sustained competitive advantage. This assumption about all data being strategic may lead to confusion about what constitutes normal inputs versus what constitutes inputs that make a real difference in the market.

The real issue lies in defining what types of data are truly significant, i.e., what data sets are not only significant but also critical to competition. Without such a distinction, any effort to address data-driven markets may be either too heavy-handed or too little. The distinction between recognizing the importance of data and recognizing its limitations makes it difficult

[10] Catherine Tucker, Digital Data, Platforms and Competition Policy, NAT'L BUREAU OF ECON. RSCH. (2019).
[11] Jared Kaplan et al., Scaling Laws for Neural Language Models (2020); Alon Halevy, Peter Norvig & Fernando Pereira, The Unreasonable Effectiveness of Data, 24 IEEE INTELLIGENT SYS. 8 (2009).
[12] ANDREW MCAFEE & ERIK BRYNJOLFSSON, MACHINE, PLATFORM, CROWD (2017).
[13] Jacques Crémer, Yves-Alexandre de Montjoye & Heike Schweitzer, Competition Policy for the Digital Era, EUR. COMM'N (2019); Autorité de la concurrence & Bundeskartellamt, Competition Law and Data (2016).

to apply existing competition law frameworks.

LIMITS OF THE EXISTING COMPETITION LAW FRAMEWORK IN INDIA

Although the phenomenon of data-driven markets has been accepted, the available legal instruments to tackle the challenge in India have remained conceptually weak. The Competition Act, 2002, was drafted at a time when market dominance was understood through price manipulation, output restriction, and structural dominance. Although the indicators continue to prevail in the enforcement environment, they seem to be at odds with the data-driven markets phenomenon.

One such limitation is apparent in the continued price-centric assessment.[14] Many of the most influential digital players have a business model that is predicated on a zero-price strategy.[15] In other words, many digital players offer free search services, free communication tools, or free access to platforms. In such a scenario, the assessment of harm to consumers through price increase does not seem to apply. In the absence of price increase, the harm, if any, is more likely to exist through a lack of innovation, a lack of privacy, and entry restriction through data asymmetry.[16] However, the assessment of the harm through the aforesaid parameters is still difficult.

A second, more structural gap relates to the lack of a clear doctrine of data concentration. While the Competition Commission of India (CCI) has dealt with several cases involving digital platforms in the last few years, its treatment of data-related issues appears hesitant.[17] Issues of data hoarding, favorable treatment of certain players, or the strategic withholding of data have been flagged in several cases. However, these have not yet been fully incorporated into the framework of assessing dominance or abuse. This creates a level of uncertainty, at least, regarding the potential for data-related issues being assessed.

The doctrine of refusal to deal may be a potential avenue for addressing these data-related issues.[18] However, the law as it currently stands provides little guidance. While the law

[14] ARIEL EZRACHI & MAURICE E. STUCKE, VIRTUAL COMPETITION (2016); MAURICE E. STUCKE & ALLEN P. GRUNES, BIG DATA AND COMPETITION POLICY (2016).
[15] John M. Newman, Antitrust in Zero-Price Markets, 128 HARV. L. REV. 149 (2015).
[16] OECD, CONSUMER POLICY AND THE DIGITAL ECONOMY (2019).
[17] Competition Commission of India, MARKET STUDY ON E-COMMERCE IN INDIA (2020); Competition Commission of India, TELECOM SECTOR IN INDIA: MARKET STUDY (2021).
[18] PHILLIP E. AREEDA & HERBERT HOVENKAMP, ANTITRUST LAW (Essential Facilities section).

acknowledges that the refusal of market access may, in certain cases, be considered an abuse of dominance, it does not address the issue of when a particular resource becomes sufficiently important so as to trigger the obligation of sharing. Nor does it address the issue of the level of indispensability of resources. In the absence of any clear guidelines, the application of the doctrine of refusal to deal may be risk averse.

This doctrinal ambiguity is further compounded by the relative silence of Indian case law on the issue of data as a competitive asset.[19] Even in platform-related disputes, such as those involving food delivery aggregators or digital marketplaces, issues of data access and control have been relegated to the background as secondary considerations rather than as primary determinants of market power. Accordingly, the regulatory framework is such that the significance of data is recognized on paper but is yet to be effectively operationalized.

These limitations indicate that the current framework is less inadequate than incomplete. Accordingly, the current framework recognizes the possibility of exclusionary conduct but fails to provide the necessary conceptual tools to recognize exclusionary conduct when it takes new forms. Any effort to address the issue of data-driven dominance must begin with the determination of when and under what circumstances data can be said to limit competition.

THE ESSENTIAL FACILITIES DOCTRINE AND ITS LIMITS IN DATA MARKETS

One possible way forward in dealing with exclusionary control in critical resources is via the Essential Facilities Doctrine (EFD). The traditional approach to EFD is where a dominant firm possesses an essential resource that is indispensable to its rivals to compete in a given market.[20] In this case, denial of access to the resource can be considered an abuse of dominance, particularly where the resource cannot be reasonably replicated.

The doctrine appears to be particularly applicable in data-driven markets. In this regard, where data is considered an essential resource to compete in a given market, denial of access to data can be considered to limit entry into a given market. However, the applicability of EFD in data-driven markets has been subject to considerable skepticism, and rightly so.

[19] Manish Agarwal, INDIAN COMPETITION LAW AND DIGITAL MARKETS (2022).
[20] MCI Commc'ns Corp. v. AT&T Co., 708 F.2d 1081 (7th Cir. 1983); Oscar Bronner GmbH & Co. KG v. Mediaprint Zeitungs- und Zeitschriftenverlag GmbH & Co. KG, Case C-7/97, 1998 E.C.R. I-7791 (EU).

The key issue with EFD in data-driven markets is that data is a non-rival resource.[21] In this case, several firms can access similar data at any given time. In this regard, it is questionable whether data can be considered indispensable to competing in a given market. In most cases, it is possible for rival firms to access data even where it requires considerable investment to access data.

Relatedly, it has also been argued that data can be substitutable.[22] That is to say, different datasets can perform similar functions in serving a market where user behavior is not particularly specialized. Relatedly, if data is substitutable, then perhaps denying access to a particular data set does not necessarily remove competition. In this sense, then, perhaps extending EFD to data is not necessary or justified.

A further concern with EFD is that it can undermine incentives to innovate. That is to say, datasharing obligations can create disincentives to invest in data collection and analysis because the benefits of doing so must be shared with others. This can create an odd situation in which datasharing obligations can undermine efficiencies and innovation, in a way that competition law traditionally seeks to avoid.

These are all important arguments, and any attempt to engage with EFD in data markets needs to take these arguments seriously. However, they are also arguments that tend to be very general in nature. In seeking to address data as a category, they perhaps miss the point that not all data is equal. While some data may be replicable or substitutable in nature, other data sets are perhaps not so easily replicable.[23]

So, the more useful question is not if data can ever be an essential facility, but when it might become one. Instead of completely rejecting the doctrine or using it in all situations, it is better to find a smaller group of situations where intervention may be necessary.[24] To do this, we need to stop talking about abstract ideas and start looking more closely at the specific features of highvalue datasets, especially in markets driven by AI.

[21] Hal R. Varian, Big Data: New Tricks for Econometrics, 28 J. ECON. PERSPS. 3 (2014); Charles I. Jones & Christopher Tonetti, Nonrivalry and the Economics of Data, 110 AM. ECON. REV. 2819 (2020)
[22] Anja Lambrecht & Catherine Tucker, Can Big Data Protect a Firm from Competition?, 8 COMPETITION POL'Y INT'L 1 (2017).
[23] Lambrecht & Tucker, "Can Big Data Protect a Firm?" (2017); Mayer-Schönberger & Ramge, Reinventing Capitalism in the Age of Big Data (2018)
[24] HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY: THE LAW OF COMPETITION AND ITS PRACTICE (6th ed. 2020).

WHEN DOES DATA BECOME "ESSENTIAL"? A DOCTRINAL TEST

To make the Essential Facilities Doctrine work for data we need a framework. If we treat all data as essential it can lead to problems. Give too much power to some companies. On the hand if we completely ignore the idea, we will not see what is really going on in some digital markets. What we really need is a test that can identify cases without making rules. This paper suggests a fourpart framework to decide when a dataset can be considered essential in competition law.

First, we need to look at if the data can be replaced. If a dataset can be recreated in an amount of time and for a reasonable cost, it cannot be called essential.[25] We should not just look at if it is technically possible. At the real challenges that new companies face. For example, in markets that use intelligence historical data collected over years may be very hard to replicate even if similar data can be gathered in theory.

Second, we need to consider scale. Some datasets are valuable not just because they exist but also, because of their size and diversity.[26] Machine learning systems in particular work better with amounts of data. A smaller dataset may not be less effective. It may not be good enough. When a system relies on access to a lot of data not having that data can create a disadvantage that's hard to overcome.

Third, we need to think about market foreclosure. The key question is whether not access to the dataset having significantly harms a company's ability to compete in the market.[27] This does not mean that all competition is eliminated. It means more than causing some problems. Not having the dataset must put companies at a disadvantage.[28]

Fourth, we need to look at centrality. The dataset must influence how the service works.[29] This is especially important in intelligence systems, where data is not just an input but a key factor

[25] Jacques Crémer, Yves-Alexandre de Montjoye & Heike Schweitzer, Competition Policy for the Digital Era, EUR.
COMM'N (2019); OECD, Big Data: Bringing Competition Policy to the Digital Era (2016).
[26] Erik Brynjolfsson & Kristina McElheran, The Rapid Adoption of Data-Driven Decision-Making, 33 AM. ECON.
REV. PAPERS & PROC. 133 (2016).
[27] RICHARD WHISH & DAVID BAILEY, COMPETITION LAW (9th ed. 2018).
[28] MASSIMO MOTTA & MARTIN PEITZ, BIG TECH AND THE DIGITAL ECONOMY: THE MOLIGOPOLY SCENARIO(2020).
[29] AJAY AGRAWAL, JOSHUA GANS & AVI GOLDFARB, PREDICTION MACHINES: THE SIMPLE ECONOMICS OF ARTIFICIAL INTELLIGENCE (2018).

in determining output quality.[30] If a service's performance depends on the dataset, controlling that data becomes crucial for competition.

This framework shows the reason for considering only specific data types. In search engine markets, for example, large-scale query data might meet all these four criteria: it might be hard to replicate, might gain value from scale, might be vital to viability, and might be vital to service performance. In intelligence systems, training data might meet all these four criteria if there is diversity in high-quality inputs. However, more generic data might not meet these criteria.

The value in this approach is in its limitation. This approach does not broadly apply the Essential Facilities Doctrine. This approach limits the Essential Facilities Doctrine to specific situations. This approach strikes a balance in preventing harmful practices and promoting innovation. The data has to be important by satisfying these four criteria. This approach directs the debate on data and competition law. The Essential Facilities Doctrine has to be limited to data that is deemed essential. This framework identifies when data is deemed essential.

POLICY DESIGN AND INSTITUTIONAL REFORM IN INDIA

Of course, we need to be aware that there might be some data sets that require regulatory intervention in particular circumstances; that's the starting point. The real question is how we take that into account in a way that informs Indian policy.

The first step towards that is to stop imposing data sharing as a blanket policy.[31] Such a policy could be damaging to the country in terms of investment and could create administrative headaches. Also, this doesn't really consider the various types of data and the market we are in; any policy we decide on needs to be precise.

In this respect, the Competition Commission of India would have to take a case-by-case approach to this problem, with the help of a structured analytical framework similar to the one discussed earlier.[32] Instead of trying to vaguely define what exactly "essential data" means, the CCI would have to look at the data set in terms of what its specific characteristics are and what

[30] Hal R. Varian, Artificial Intelligence, Economics, and Industrial Organization (2019).
[31] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use Across Societies(2019); European Commission, A European Strategy for Data (2020).
[32] Jason Furman et al., Unlocking Digital Competition: Report of the Digital Competition Expert Panel (2019).

type of environment it exists in.[33]

If the CCI does have to get involved in this case, then instead of enforcing free access to data, the focus would have to be on promoting competition in a less intrusive manner. Data portability and interoperability, along with API access, would be ways to promote competition in this respect.[34] These are being considered in other countries as well and would be effective in this case.[35]

However, the CCI would need to focus on increasing its capabilities in order to have a favorable regulatory environment in data-driven markets. It is impossible to fully understand the technical aspects of AI and data ecosystems from a legal perspective.[36] To improve its research and analytical procedures, the CCI would therefore need to collaborate with data scientists and specialists in digital markets. If not, it would be challenging to implement even a good system. Lastly, a gradual approach to regulations can be implemented, whereby the CCI may consider publishing guidelines or whitepapers outlining their opinions on specific data-related issues rather than directly imposing regulatory systems on organizations. Even though the regulator can change its mind over time, this would give market players more clarity.

CONCLUSION

It is in this context that one is of the opinion that the increase in the importance of data in relation to determining parameters of competitiveness is likely to pose both opportunities and challenges to competition law. It is likely to require a new perspective in relation to our understanding of assessing market power in a world of digital markets and AI technology, and yet, it is likely to require that we do not fall into the trap of relying upon existing principles in this context.

The paper has sought to further advance the case for a more nuanced approach to assessing the notion of data as an essential facility, rather than accepting that all data is, in fact, essential, and to further advance a set of specific circumstances in which data might be of concern, rather than seeking to accept a debate in which sweeping statements are often made.

[33] Australian Competition & Consumer Commission, Digital Platforms Inquiry Final Report (2019).
[34] General Data Protection Regulation (EU) 2016/679 art. 20 (EU).
[35] OECD, Data Interoperability: Implications for Competition and Innovation (2021).
[36] OECD, OECD Framework for the Classification of AI Systems (2021).

The paper also talks about the perils of having too little and too much in the realm of regulations. The problem is that, on one hand, having too many regulations in the realm of access requirements may result in new ideas never being allowed to surface, but on the other hand, having too little in the realm of existing barriers to entry may result in the existing power structures being further entrenched. The challenge, therefore, is to find that sweet spot which keeps the energy vibrant without losing sight of new ways of keeping people out.

In the end, it is not new ideas on which the law of competition in a data-driven market will be based. It will be based on how we implement the ideas we already have. This means that in India, there needs to be a system in place to address the technical realities.