European Competition Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/recj20

Charting a way forward for the use of data science in competition enforcement and platform regulation

Inge Graef, Ulrich Laitenberger & Jens Prüfer

To cite this article: Inge Graef, Ulrich Laitenberger & Jens Prüfer (13 Nov 2024): Charting a way forward for the use of data science in competition enforcement and platform regulation, European Competition Journal, DOI: 10.1080/17441056.2024.2428034

To link to this article: https://doi.org/10.1080/17441056.2024.2428034

Published online: 13 Nov 2024.

EUROPEAN COMPETITION JOURNAL
https://doi.org/10.1080/17441056.2024.2428034

Routledge
Taylor & Francis Group

Charting a way forward for the use of data science in competition enforcement and platform regulation*

Inge Graef [a,b,c], Ulrich Laitenberger[d,e] and Jens Prüfer[f,g,h]

[a]Associate Professor of Competition Law at Tilburg University; [b]Tilburg Institute for Law, Technology and Society (TILT); [c]Tilburg Law and Economics Center (TILEC); [d]Associate Professor of Economics at Tilburg University; [e]Academic Lead of the Information Systems Section; [f]Director of the Tilburg Law and Economics Center (TILEC); [g]Professor of Economics at the University of East Anglia's School of Economics; [h]Deputy Director of the Centre for Competition Policy (CCP)

ABSTRACT

Competition authorities and other regulators already rely on data science to monitor markets and check compliance with applicable rules. The Digital Markets Act and the Digital Services Act have made the use of data science by regulators even more relevant, given the scale and complexity of the oversight required. The existing use of data science shows that regulators are able to adapt their organizations and processes to realize the potential of using technology tools in their activities. At the same time, the use of data science poses challenges in terms of data reliability and individual privacy. Beyond involvement in specific investigations, data science also has the potential to reform the work of regulators by moving towards a more proactive form of enforcement. The sharing of data science expertise and tools between regulators deserves to be further facilitated to increase collaboration and share resources in monitoring the platform economy.

ARTICLE HISTORY Received 24 July 2024; Accepted 4 October 2024

KEYWORDS Artificial intelligence; digital markets act; digital services act; platform markets; content moderation; self-preferencing

I. Introduction

Together with large amounts of data, artificial intelligence (AI) and especially machine-learning (ML) algorithms offer the key ingredients to data science. AI and data technologies are increasingly used by

CONTACT Inge Graef i.graef@tilburguniversity.edu

*This paper is based on work that the authors have done as part of the expert group for the EU Observatory on the Online Platform Economy. All contents, opinions and errors are our own and do not reflect the position of the European Commission.

competition authorities and other regulators to support their monitoring tasks. Particularly in current complex platform markets, the systematic collection and analysis of data can give more insight into the impact of practices on our economy and society. The calls for greater transparency and accountability in the platform economy have led to the adoption of the Digital Markets Act (DMA)[1] and the Digital Services Act (DSA),[2] which introduce a range of obligations and prohibitions. However, these new rules will only reach their objective with proper monitoring and enforcement. Data science can be a powerful tool to support regulators in this task.

Several regulators, including the European Commission,[3] have started to invest in the use of AI and data science tools for monitoring and enforcement. Distinguishing themselves from traditional statistics and econometrics, these methods use algorithmic models and treat the underlying data-generating patterns as unknown in order to discover complex structures that were not specified in advance. Where conventional statistics is deductive, data science is inductive. These inductive methods facilitate the automated collection of information, especially on, but not restricted to, the Internet. Via text analysis, computers can learn to understand the meaning of words, relate them to each other, and analyse them at scales that otherwise would require the help of hordes of research assistants. The new techniques and technologies also allow to use many more (unstructured) real-time data sources to conduct investigations that would not have been possible otherwise, for instance by using sensor data from mobile devices.[4] By making reliance on subjective and self-reported surveys largely unnecessary and substituting these sources with objective data on revealed preferences, they improve the accuracy, robustness and, hence, the value of regulatory oversight.

Against this background, the paper sheds light on how data science technologies and data science expertise can be used to monitor markets, check compliance with laws, and support enforcement in the

[1]Regulation (EU) 2022/1925 of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) OJ [2022] L 265/1.

[2]Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) OJ [2022] L 277/1.

[3]In particular, the European Commission has established the European Centre for Algorithmic Transparency (ECAT) that "contributes with scientific and technical expertise to the European Commission's exclusive supervisory and enforcement role of the systemic obligations on Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) provided for under the DSA". See <https://algorithmic-transparency.ec.europa.eu/index_en>.

[4]Joshua Blumenstock, Gabriel Cadamuro and Robert On, 'Predicting Poverty and Wealth from Mobile Phone Metadata' (2015) 350(6264) Science 1073.

platform economy. Its aim is to help frame the debate and set an agenda by exploring future directions to use data science in competition enforcement and platform regulation. Section II reflects on the current use of data science technologies. Section III analyses some promises and challenges of the future uptake of existing and new data science tools. And Section IV discusses possibilities for integrating data science expertise into regulation and policy by drawing lessons from existing experience across regulators.

II. Overview of existing data science and technology tools for compliance and enforcement in the platform economy

In the context of competition law and platform regulation, the adoption of data science and technological tools offers increasing potential for ensuring compliance and effective enforcement.[5] In this light, the objectives of this section are twofold: firstly, to catalogue the data science and technological tools that are presently available to compliance and enforcement activities; and secondly, to provide an analysis of their current applications. After an initial discussion on the typical requirements for quantitative analyses, this section will explore data collection mechanisms such as web scraping technologies and Application Programming Interfaces (APIs), as well as data processing and analytical tools that employ sophisticated algorithms. We finish by discussing the potential and challenges of these tools for specific upcoming regulatory and enforcement tasks.

A. Understanding typical needs for the quantitative analysis of platform behaviour

Measuring and characterizing behaviour of platforms in digital markets has certain advantages but also challenges as compared to the study of traditional markets. The main advantage is that digital platforms document all transactions and interactions with and between their users in great detail. Fostered by the low cost of storing data and emerging analysing techniques, platforms have an incentive to collect detailed data as it can help them to optimize their business. However, for a third-party

[5]For an analysis of the use of technology for enforcement in consumer law, see Christina Riefa and Liz Coll, 'The Transformative Potential of EnfTech in Consumer Law' (2024) <https://www.enftech.org/report>. For an overview of the use of computational methods by competition authorities, see Thibault Schrepel and Teodora Groza, 'Computational Antitrust Within Agencies: 3rd Annual Report' (2024) 53 (4) Stanford Computational Antitrust <https://doi.org/10.2139/ssrn.4861858>.

that is interested in studying the behaviour of a specific platform, several challenges arise.

Data collection: While the platform may collect vast amounts of data, this data stays often inside the business and cannot be accessed by third parties. While regulators can often use their formal powers to request information, data collection can be a problem if a regulator aims at screening platform behaviour to identify cases for a proper investigation.

Data processing and analysis: Even if data is available or extractable for instance by web scraping, its processing and analysis can be quite complex. Platform data often comes unstructured, dispersed across various information systems, and includes relevant textual content. Traditional approaches, often labour-intensive and manual, increasingly need to be replaced by automated systems anchored in machine learning algorithms and computational models.

Data visualization: Finally, the relationship of the users among themselves and with the platform has a complex structure, requiring techniques that make this knowledge accessible to those with less technical expertise including lawyers and judges.

We will discuss a few examples of techniques that mitigate these challenges.

B. Data collection

The first challenge for an empirical investigation is the retrieval or collection of data. In this regard, competition authorities and other regulators hold formal powers to request information from firms in order to conduct investigations and monitor compliance with legal requirements.[6] In addition, legislative instruments sometimes require firms to make certain data available to the public. An example is the requirement imposed on providers of very large online platforms and very large online search engines in the DSA that display advertising in their online interfaces "to compile and make publicly in a specific sector of the online interface, through a searchable and reliable tool that allows multicriteria queries, and through application programming interfaces, a repository" containing the advertisements displayed in the platform.[7]

[6]For instance, the power of the European Commission to request information in the context of competition investigations is laid down in Article 18 of Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty (Regulation 1/2003) OJ [2003] L 1/1. The relevant provisions in the DMA and DSA are, respectively, Article 21 of the DMA and Article 67 of the DSA.

[7]DSA, art. 39.

However, beyond these contexts, data collection can be challenging. Surveys on academic studies using data collection approaches include the work of Edelman from 2012 and of Boegershausen and others from 2022.[8]

Web scraping is a common practice to collect data about a platform. It refers to the use of technology tools for gathering data from webpages in an efficient and automated manner, rendering the data in a more structured and easier-to-use format.[9] Data scraping allows for the transformation of unstructured data into various formats for later manipulation and analysis. Different tools exist for scraping data in different formats (for example, tools for parsing and scraping data in HTML format, PDFs, etc.), with varying level of complexity for their users. Web scraping usually allows to collect information such as the assortment in terms of product, service or content offers; their characteristics such as prices, public consumer feedback (ratings and reviews); their positioning (i.e. rankings, recommendations likelihoods) as well as their popularity.[10] Web scraping may also be used to conduct "sock puppet" experiments, that is, creating bots to engage with the platform and see its behaviour. For example, one can create an account as a minor, and use web scraping bots to examine whether a platform serves ads. Web scraping is less useful in a context where the user interaction happens in a non-public way, for instance if prices and conditions are negotiated bilaterally between suppliers and buyers. If the data on the platform is very dispersed (e.g. many products, users, etc.), collection may take a lot of time and resources. Also, usually only current data is available, which renders analyses in retrospect impossible.

Application Programming Interfaces (APIs) facilitate data retrieval from software applications. These are typically offered by the platform themselves. APIs can provide access to non-public internet environments, including those requiring login authentication, as the data is shared "directly through the back-end of the social media service to which the data belong".[11] Examples are large social media platforms

[8]Benjamin Edelman, 'Using Internet Data for Economic Research' (2012) 26(2) Journal of Economic Perspectives 189 and Johannes Boegershausen and others, 'Fields of Gold: Scraping Web Data for Marketing Insights' (2022) 86(5) Journal of Marketing 1.

[9]Osmar Castrillo-Fernández, 'Web Scraping: Applications and Tools' (2015), European Public Sector Information Platform Topic Report No. 2015/10 and Harshit Nigam and Prantik Biswas, 'Web Scraping: From Tools to Related Legislation and Implementation Using Python' in Jennifer S Raj and others (eds), *Innovative Data Communication Technologies and Application* (Springer 2021) 149–64.

[10]Johannes Boegershausen and others, 'Fields of Gold: Scraping Web Data for Marketing Insights' (2022) 86(5) Journal of Marketing 1.

[11]Irvin Dongo and others, 'A Qualitative and Quantitative Comparison Between Web Scraping and API Methods for Twitter Credibility Analysis' (2021) 17(6) International Journal of Web Information Systems 580.

such as Meta, which provide third parties and researchers with access to data for their analyses to some extent.[12] Sometimes APIs exist for accessing platform data which are developed and commercialized by third parties. While these usually do not provide other data than that is publicly available on the original website, they may facilitate greatly the collection of data and often also allow for the retrieval of historical data. Drawbacks of first-party APIs are that they are restricted to what the platform perceives useful to share with other stakeholders. Furthermore, there is a tendency that platform-initiated APIs become less available, such that they cannot be employed for long-term investigations.[13] Third-party APIs suffer from the same data restrictions as web scraping. Furthermore, they may be selective in the sense that they merely collect data which is of interest for other market participants, but not necessarily for researchers or regulators, requiring customized web scraping solutions.[14] Both types of APIs are often restricted to data types that are less likely to fall within the scope of data protection laws.[15]

C. Data processing and analysis

Regulators and enforcers may want to analyse data for the purpose of identifying behavioural patterns. As examples of problematic patterns, one can think of anticompetitive pricing (collusion, predatory behaviour), biased recommendations (such as self-preferencing), unintentional or intentional discrimination and a lack of content moderation (covering issues like hate speech, child protection, etc.). The challenge is in having adequate methods at hand that take the specific nature of platform-related data into account, while achieving a high degree of comprehensibility and interpretability. We discuss machine learning (ML) techniques, natural language processing and network analysis here, as these have gained popularity recently. Beyond this, we note that some less typical

[12]Stine Lomborg and Anja Bechmann, 'Using APIs for Data Collection on Social Media' (2014) 30(4) The Information Society 256. Note that Article 40(4) of the DSA requires very large online platforms and very large online search engines to give vetted researchers access to data to conduct research into systemic risks.

[13]Jessamy Perriam, Andreas Birkbak and Andy Freeman, 'Digital Methods in a Post-API Environment' (2020) 23(3) International Journal of Social Research Methodology 277.

[14]Daniel Glez-Peña and others, 'Web Scraping Technologies in an API World' (2014) 15(5) Briefings in Bioinformatics 788.

[15]Irvin Dongo, Yudith Cardinale and Ana Aguilera, 'Credibility Analysis for Available Information Sources on the Web: A Review and a Contribution' (2019) 4th International Conference on System Reliability and Safety (ICSRS) 116 and Irvin Dongo and others, 'Web Scraping versus Twitter API: A Comparison for a Credibility Analysis' (2020) Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services 263.

approaches that we do not discuss, like causal analysis, may also be relevant for the enforcement context.

Before discussing techniques to deal with textual data, we will start off with an overview of **ML techniques** that can be used to identify patterns in large datasets. The resulting models can be used for classification or making predictions.[16] We will discuss some methods here briefly.

- **Supervised machine learning** techniques may be employed in a situation when training data is available. This is, for example, the case when for a subset of the data, it has already been identified whether there is a problem or not. Supervised ML algorithms can learn from this data, such that they are trained to predict if cases are problematic, either for a binary ("classification algorithm") or a continuous outcome ("regression algorithm"). Of particular interest for regulators and enforcers are decision trees. These constitute sequential models that logically combine a simple test that compares a numeric attribute against a threshold value or a nominal attribute against a set of possible values. Decision trees are popular as they are more comprehensible and interpretable than other techniques, as reflected by the wide use in financial firms.[17] An extended form of tree classifiers is the **random forest** approach, which is described as one of the best performing methods. **Gradient boosting** and **k-nearest neighbours (KNN)** are techniques to improve the accuracy and predictive capabilities of ML models. However, especially the latter mentioned method lacks transparency regarding how specific variables affect the outcome. **Support vector machines (SVM)** is another supervised ML technique for classification tasks that has been used for fraud detection in the past.
- In contrast to supervised machine learning methods, **unsupervised machine learning** does not focus on training the machine to predict specific outcomes. Instead, its objective is to describe the entities involved and the relationships between them. This type of machine learning has multiple applications in the field of regulation and enforcement. For instance, algorithms that perform **clustering analysis** use

[16]Speech by Stefan Hunt (on behalf of the UK Financial Conduct Authority), 'From Maps to Apps: the Power of Machine LEarning and Artificial Intelligence for Regulators' (2017) Beesley Lecture Series on Regulatory Economics 19 October 2017 <https://www.fca.org.uk/publication/documents/from-maps-to-apps.pdf> and Mohamed Alloghani and others, 'A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science' in Michael W Berry, Azlinah Mohamed and Bee Wah Yap (eds), *Supervised and Unsupervised Learning for Data Science* (Springer 2019) 3–21.

[17]Sotiris B Kotsiantis, 'Decision Trees: A Recent Overview' (2013) 39 Artificial Intelligence Review 261.

unstructured data to categorize different market actors – such as firms, consumers, and traders – based on their similar behaviours. The UK Financial Conduct Authority (FCA) employed such clustering algorithms to discern patterns among consumers who consistently overdraft their bank accounts. The US Commodity Futures Trading Commission also used these algorithms to categorize traders by their activities.

Some techniques can fall under either the "supervised" or "unsupervised" categories, depending on how they are deployed and the type of data they are trained on. An example is **neural networks**, which are a subset of machine learning known as "deep learning". They are designed to mimic the human brain and consist of interconnected nodes organized into layers. These networks are commonly utilized in various tasks, including image recognition, natural language processing, and predictive analytics. For instance, Xu and others made use of a natural language processing neural network to identify illegal trading of ivory and pangolin on Twitter.[18]

Natural language processing (NLP) techniques can be used to process and analyse textual information in a supervised and unsupervised fashion, and be combined with the previously described ML methods. Examples for documented applications can be found in accounting, auditing, financial and judicial fields which extract information from company-issued reports or court documents to classify a behaviour or to predict a certain outcome. An area of NLP is **sentiment analysis** which aims at identifying, extracting, and organizing sentiments contained in text data collected from social networks, blogs, and other sources.[19] Finally, **topic modelling** allows for the identification of patterns and themes in a corpus of texts. It has been especially useful for enforcement purposes and more generally for applications involving data sourced from social media platforms, which include texts that may be unstructured or unlabelled, and involve discourse between multiple users.[20]

[18]Qing Xu and others, 'Use of Machine Learning to Detect Wildlife Product Promotion and Sales on Twitter' (2019) 2 Frontiers in Big Data 1.

[19]Zhongnan Zhao, Wenjing Liu and Kun Wang, 'Research on Sentiment Analysis Method of Opinion Mining Based on Multi-Model Fusion Transfer Learning' (2023) 10 Journal of Big Data 1.

[20]Birgit Kirsch and others, 'Robust End-User-Driven Social Media Monitoring for Law Enforcement and Emergency Monitoring' in Georgios Leventakis and MR Haberfeld (eds), *Community-Oriented Policing and Technological Innovations* (Springer 2018) 29–36 and Murimo Bethel Mutanga and Abdultaofeek Abayomi, 'Tweeting on COVID-19 Pandemic in South Africa: LDA-Based Topic Modelling Approach' (2022) 14(1) African Journal of Science, Technology, Innovation and Development 163.

We conclude our overview with techniques and tools for **network analysis**. These can be used to leverage relational data, often sourced from APIs, to illustrate connections between various entities, such as users on a platform. Whether showing social links among users or the type of information exchanged between them, network analysis employs nodes and edges to offer a robust framework for understanding and visualizing complex relationships.

D. Data visualization

Besides tools for data extraction and processing, visualization tools are also valuable for presenting crucial information in formats that aid regulatory and enforcement activities. For instance, **interactive dashboards** facilitate real-time data scrutiny, helping regulators to quickly spot patterns, trends, or irregularities. These dashboards can integrate features like geospatial mapping to overlay regulatory information on geographical maps, thereby illustrating compliance levels, violations, and enforcement actions in a spatial context. In addition, visualization may be particularly important to communicate the results of analyses to non-technical audiences such as lawyers and judges.

E. Potential and challenges

Data science tools can assist in monitoring the behaviour of platforms and provide an initial quantification of undesirable outcomes. This relevance is heightened by new legal instruments like the DMA and DSA, which impose obligations and prohibitions on specific types of online businesses, referred to as "gatekeepers" and "very large online platforms or search engines", respectively. For some of these regulatory mandates, preliminary analyses using publicly available data are feasible and could support the enforcement process. Yet, for more complex issues, direct collaboration with or the use of formal information requests against the scrutinized platform may be essential. To elaborate, we will explore specific examples.

Example 1: bias in recommender systems: Search engines and e-commerce platforms help consumers discover a variety of products, services, and content from multiple suppliers. These platforms deploy intricate algorithms that guide consumer choices through recommendations, thereby easing their search process.[21] However, concerns arise that

[21] Matthias Hunold, Ulrich Laitenberger and Guillaume Thébaudin, 'Bye-Box: An Analysis of Non-Promotion on the Amazon Market Place' (2022) CRED Discussion Paper 2022 N°4.

platforms take "unfair" factors into account when providing these recommendations.[22] This happens, for instance, when these platforms favour their own products or those of their subsidiaries (self-preferencing), a specific type of producer[23] or unfairly penalize third-party offers based on behaviour external to the platform.[24] Data science tools can offer initial insights into whether such problematic factors influence these algorithms. Enforcers may use web scraping methods to collect data on prices and product features, and recommendations given by the platform. They can employ machine learning techniques to analyse how these factors influence recommendations. However, it is crucial to note that a meaningful study of recommendations given by a platform requires a solid understanding of how actual consumers make search requests on the platform.[25] Enforcers need to develop a methodology that is representative of user behaviour. Furthermore, investigators need to be aware that other factors that are not observable for them via public data may be correlated with problematic or "unfair" factors (endogeneity), requiring additional data. Reimers and Waldfogel illustrate different approaches by simulation as well as by using data from platforms that allow producing estimates of platform bias and its welfare cost.[26]

Example 2: content moderation: Social media platforms act as focal points for user-generated content, attracting both users and advertisers. Yet, they face challenges such as misinformation, hate speech, and intellectual property violations. Platforms are subject to regulatory obligations to moderate problematic content and they also have their own interest to do so as advertisers are concerned with "brand safety". This is a term that

[22]Lukas Jürgensmeier and Bernd Skiera, 'Measuring Fair Competition on Digital Platforms' (2023) SSRN Working Paper <https://ssrn.com/abstract=4393726>.

[23]Luis Aguiar, Joel Waldfogel and Sarah Waldfogel, 'Playlisting Favorites: Measuring Platform Bias in the Music Industry' (2021) 78 International Journal of Industrial Organization 102765.

[24]Morgane Cure and others, 'Vertical Integration of Platforms and Product Prominence' (2022) 20 Quantitative Marketing and Economics 353; Abhishek Dash and others, 'When the Umpire is also a Player: Bias in Private Label Product Recommendations on E-commerce Marketplaces' (2021) Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 873; Devesh Raval, 'Steering in One Click: Platform Self-Preferencing in the Amazon Buy Box' (2023) Working Paper <https://deveshraval.github.io/buyBox.pdf>; Xuan Teng, 'Self-Preferencing, Quality Provision, and Welfare in Mobile Application Markets' (2022) CESifo Working Paper, No. 10042, Center for Economic Studies and ifo Institute (CESifo) Munich <https://www.econstor.eu/bitstream/10419/267275/1/cesifo1_wp10042.pdf>; Matthias Hunold, Reinhold Kesler and Ulrich Laitenberger, 'Rankings of Online Travel Agents, Channel Pricing, and Consumer Protection' (2020) 39(1) Marketing Science 92.

[25]Chiara Farronato, Andrey Fradkin and Alexander MacKay, 'Self-Preferencing at Amazon: Evidence from Search Rankings' (2023) 113 AEA Papers and Proceedings 239.

[26]Imke Reimers and Joel Waldfogel, 'A Framework for Detection, Measurement, and Welfare Analysis of Platform Bias' (2023) National Bureau of Economic Research Working Paper 31766 <https://www.nber.org/system/files/working_papers/w31766/w31766.pdf>.

refers to the suitability of content in the context of a brand's image.[27] Although platforms have both intrinsic and extrinsic motives to moderate content – stemming from regulatory obligations and advertiser concerns about brand safety – these interests are counterbalanced by their aim to maintain user engagement. Regulators aiming to evaluate the effectiveness of content moderation algorithms can utilize Natural Language Processing (NLP) techniques. These techniques can analyse content collected through web scraping or API access.[28] However, this approach has limitations, such as restricted access to the complete content on the platform, and the nuanced identification of what constitutes "harmful" content. For instance, techniques such as sentiment analysis may only give a probabilistic assessment of how problematic certain content is.

The examples discussed offer initial insights into how competition authorities and other regulators can use data science tools for a preliminary quantitative assessment of the prevalence of certain problematic behaviours. This can be accomplished through the collection of data and the application of well-established analytical techniques. However, it is important to note that these data collection methods typically yield only a representative sample, rather than a complete picture. Additionally, due to the probabilistic nature of most analytical techniques and the unavailability of certain data, absolute conclusions are generally unattainable. This also implies that these techniques should be used to support and not substitute the analysis done by human investigators at regulatory agencies, whose decisions need to meet a certain standard of proof to be upheld in courts. By recognizing both the strengths and limitations of data science tools in these specific situations, competition authorities and other regulators can more effectively enforce compliance with laws governing the platform economy.

[27]Yi Liu, T Pinar Yildirim and Z John Zhang, 'Implications of Revenue Models and Technology for Content Moderation Strategies' (2022) 41(4) Marketing Science 831 and Leonardo Madio and Martin Quinn, 'Content Moderation and Advertising in Social Media Platforms' (2023) Marco Fanno Working Papers – 297 <https://www.economia.unipd.it/sites/economia.unipd.it/files/20230297.pdf>.

[28]Raphaela Andres and Olga Slivko, 'Combating Online Hate Speech: The Impact of Legislation on Twitter' (2021) ZEW Discussion Papers 21–103 <https://www.econstor.eu/bitstream/10419/248857/1/1785234617.pdf>; Rafael Jiménez Durán, Karsten Müller and Carlo Schwarz, 'The Effect of Content Moderation on Online and Offline Hate: Evidence from Germany's NetzDG' (2023) SSRN Working Paper <https://ssrn.com/abstract=4230296>; Karsten Müller and Carlo Schwarz, 'From Hashtag to Hate Crime: Twitter and Antiminority Sentiment' (2023) 15(3) American Economic Journal: Applied Economics 270; Karsten Müller and Carlo Schwarz, 'The Effects of Online Content Moderation: Evidence from President Trump's Account Deletion' (2023) SSRN Working Paper <https://ssrn.com/abstract=4296306>.

III. Promises and challenges of the future uptake of existing and new data science tools

The previous section has shown that the realm of data science tools is extensive, offering a diverse array of methods and technologies, each with their own set of advantages, drawbacks, and relevance to particular situations. Choosing the most appropriate tool for a given regulatory issue demands a thorough comprehension of the issue at hand, the data available, the goals to be achieved, and the technical skills needed. Additionally, distinct regulatory fields in the platform economy may present unique data-related challenges, legal limitations, or analytical aims that call for specialized approaches. This makes it essential to meticulously assess the available tools to gauge their aptness for tackling particular regulatory issues. Building on this analysis, this section will discuss a couple of promises and challenges of the future uptake of existing and new data science tools.

A. Democratization of investigations as a promise of data science

Given that data science methods are usually freely available and relatively easy to learn, data science techniques thereby contribute to a "democratization" of empirical, data-based investigation tools, where regulatory agencies with fewer resources may be capable of doing similar investigations as better endowed agencies and, especially, as regulated tech firms from resource-rich countries. This process of democratization also offers opportunities for the public, researchers and NGOs to investigate the conduct of tech firms.

However, data science methods cannot substitute human creativity and research design skills. According to Agrawal, Gans & Goldfarb, AI algorithms are better than humans at factoring in complex interactions among different indicators if enough data is available.[29] If this condition does not hold, however, humans are often better than machines in understanding the data. For regulatory oversight, data science methods appear to be especially well suited for first, inductive analyses that guide further investigation efforts. This occurs, for instance, by pointing regulatory agencies at relevant correlations and helping them to design better regulatory sandboxes or to run their own small – or medium scale experiments (A/B-testing) to reveal behaviour that was difficult to detect previously.

[29] Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press 2018).

The inductive, data-driven approach can also point investigators at the key variables of interest for a specific case or industry. This may alleviate the need for expert interviews or the use of small, unrepresentative surveys to obtain a first understanding of the driving economic forces in a given industry. Data science techniques may also reduce the risk that investigators fall victim to confirmation bias.[30] This may lead the most advanced enforcers to a habit of motivating their investigations by the results of big data analyses.

B. Other promises, but also risks

Notably, data science methods are no substitute for traditional market investigations or conventional statistics. They complement those established methodologies. A fruitful avenue is to combine data science techniques with administrative and survey data. Data science techniques have been largely applied to Internet data (often by scraping and analysing big social media datasets), including by regulators as Section II above shows. However, this approach ignores both potential selection effects that are due to differences between online (social media) users and the entire population and measurement errors that are due to the unreliability of social media data as a representative measure of social phenomena. The DSA, with its Article 40(4), may offer one way out here as researchers have the right to access raw data of very large online platforms and search engines, including several social media providers. Beyond the DSA, comparing the results of a (small) representative survey with results of (big) unrepresentative data, of which the representativeness can be assessed empirically, therefore looks like an ideal way forward for empirical investigations.[31]

Just as all technologies based on AI, data science methods come with risks. Agrawal, Gans & Goldfarb conclude their book on the consequences of AI by focusing on three trade-offs.[32] The first is productivity versus distribution. Bughin and others note: "A key challenge is that

[30] Jasmin Mahmoodi and others, 'Big Data Approaches in Social and Behavioral Science: Four Key Trade-Offs and a Call for Integration' (2017) 18 Current Opinion in Behavioral Sciences 57.

[31] As Google's Chief Economist, Hal Varian comments: "A good predictive model can be better than a randomly chosen control group, which is usually thought to be the gold standard". See Hal Varian, 'Big Data: New Tricks for Econometrics' (2014) 28(2) Journal of Economic Perspectives 3, 23. And Seth Stephens-Davidowitz notes: "[E]ven a spectacularly successful Big Data organization like Facebook sometimes makes use of […] a small survey". See Seth Stephenson-Davidowitz, *Everybody Lies—Big Data, New Data, and What the Internet can Tell us about Who We Really Are* (Harper Collins 2017) 255.

[32] Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press 2018).

adoption of AI could widen [performance and outcome] gaps between countries, companies, and workers".[33] Applied to the enforcement context, data science methods can increase the number, breadth, and speed of questions regulatory agencies can work on, increasing their productivity. But regulatory agencies who neglect technological progress or who miss the train may be at a disadvantage as some traditional methods may be dominated by data science techniques quickly. Consequently, there may be a watershed moment for *every* agency, where it either invests resources into data science methods (for these, the above-described democratization of investigation tools may kick in), or not (which saves time and costs in the short run but may come at significant risk for the quality of their investigations in the long run, especially when regulated companies are tech-savvy and data-driven).

The second trade-off underlined by Agrawal, Gans & Goldfarb is innovation versus competition.[34] On many markets, the successes of Alphabet and Meta, both of which are highly data-driven firms that embraced AI early, have shown that data-driven markets display first-mover advantages and are prone to market tipping. Moreover, traditional markets can be transformed into data-driven markets, as Alphabet, in particular, has shown (compare markets for maps or yellow pages before and after they entered). Observing the dismal fate of their competitors underlines how important it is not to fall behind on data-driven markets.[35] Data science methods could introduce a similar spiral, where those agencies that embrace them early can produce higher-quality market investigations quicker, which may have positive feedback effects on their future cases. Consequently, the quality and reputation of the leading enforcers' work can increase tremendously, giving them significant opinion leadership and influence to shape global standards of investigations.

The third trade-off, according to Agrawal, Gans & Goldfarb, is performance versus privacy.[36] Using AI successfully depends on huge

[33] Jacques Bughin and others, 'Notes from the AI Frontier: Modeling the Impact of AI on the World Economy' (2018) Discussion Paper McKinsey Global Institute <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy>.

[34] Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press 2018).

[35] Prüfer and Schottmüller provide more empirical details, rationalise dominant firms' strategies, and introduce "data-driven indirect network effects" as the source of market tipping on data-driven markets. See Jens Prüfer and Christoph Schottmüller, 'Competing with Big Data' (2021) 69(4) Journal of Industrial Economics 967.

[36] Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press 2018).

amounts of data because it is the very power of personalization of services and inference about an individual's preferences and characteristics that can be made if only sufficient data about other individuals is available. But the benefits of aggregate data easily come at individuals' costs, especially for privacy.[37] Investigating firms' conduct on markets with millions of end users by analysing big datasets with data science methods puts regulators in a similar position, with respect to those end users, as the firms providing the regulated services. Here the General Data Protection Regulation (GDPR)[38] is relevant, which regulates the processing of personal data. It should be noted that the GDPR also applies to public agencies acting to protect consumers, as it regulates the collection and analysis of their personal data. Enforcers are thus subject to the same legal standards as the firms they are investigating, possibly affecting the scope of their investigations in order to protect the privacy of individuals – particularly given the need to maintain their reputation for upholding legal rules.

C. Explainable AI

Crucially, the one-to-one translation of the three trade-offs listed by Agrawal, Gans & Goldfarb from the business to the enforcement domain is subject to further scrutiny.[39] For instance, it is unclear whether investigations using data science methods are subject to the same indirect network effects as competition on data-driven markets (which leads to market tipping and one highly dominant firm per market). By contrast, what is certainly true is that regulators need to keep up the standards of verifiability, reliability, and replicability of their work – which will also benefit them when facing the courts. However, this is particularly difficult when ML algorithms are used because, by definition, the algorithm is learning: it adapts based on feedback. Therefore, it is harder than with conventional investigation methods to reproduce predictions (read: results) based on ML. What is

[37] For an overview of the privacy literature, see Alessandro Acquisti, Curtis R Taylor and Liad Wagman, 'The Economics of Privacy' (2016) 54(2) Journal of Economic Literature 442. For a rationalization of consumers' privacy choices even if they have no exogenous taste for privacy, see Sebastian Dengler and Jens Prüfer, 'Consumers' Privacy Choices in the Era of Big Data' (2021) 130 Games and Economic Behavior 499.

[38] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ [2016] L 119/1.

[39] Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press 2018).

necessary, thus, is to make the use of algorithms in regulators' investigations more transparent. This will both strengthen trust in the new technologies and in their results when presented in court.[40]

One option to achieve transparency is to build algorithms with an internal self-evaluation or calibration stage such that the machine can test its own certainty and report back to the data scientist. One attempt in this direction is the "Automatic Statistician", which was developed at Cambridge University.[41] The tool is set up with funding from Google and helps researchers to analyse their datasets while also providing a report in a human understandable form that explains what it is doing and how certain it is about its predictions. This technology is related to a recent development within ML, "Automated Machine Learning" (AutoML). This approach tackles the fundamental problems of accountability and verifiability. Here, ML methods and hyperparameter settings are automatically selected and, thereby, reduce the necessity of handcrafted human interventions. While ML should be prevented from turning into a black box, AutoML can substantially improve performance and provide evaluations of all tested methods and specifications. Thereby, it can help non-experts to effectively and reliably apply ML techniques.

Of course, such requirements of algorithmic transparency and replicability are especially important for regulators in external relationships (including with courts) and to ensure trust and accountability towards the public. For internal procedures, such as early-warning systems that regularly scrape data about certain markets and highlight potential violations, algorithmic transparency is less crucial at the immediate stage, as the algorithms' work will only be the first step to further human investigation.

D. How to govern by technology in the future?

Section II lists several insightful examples of actual applications of data science techniques by competition authorities and other regulators. Thinking ahead (and hoping to inspire), potential future applications could include the development of privacy-enhancing technologies to be used in data exchanges in industries subject to the Data Act and the

[40] Catalina Goanta, 'Digital Market Surveillance: The Role of Automation in Consumer Protection Enforcement' in Marion Ho-Dac and Cécile Pellegrini (eds), *Governance of Artificial Intelligence in the European Union: What Place for Consumer Protection?* (Bruylant 2023) 343, 366.

[41] See <https://www.automaticstatistician.com/index/>.

European data spaces.[42] These technologies would fit with today's initiatives on digital monitoring tools, such as a PDF checker to assess digital accessibility in line with the EU's WCAG (Web Content Accessibility Guidelines) or a (national) register of algorithms in line with the AI Act.[43]

Algorithms could support regulators by automatically monitoring or even auditing platforms' recommender systems to see whether they work in line with the DMA. In the context of social media platforms, regulators could include the development of tooling that identifies groups of social media users that are prone to problematic social media use, classify their risk profile in line with the 4 levels of risk in the AI Act,[44] and flag them to social media providers. Such tools could also check whether the social media firms' own tools work well.

Another application could be to use digital sciences for online safety and transparency, given that social media and big tech firms are obliged to fight hate speech and foster online safety of all users. Ideally, a regulator is equipped with own content moderation tooling or other digital management devices to detect hate speech posts or other online content harming fundamental human rights, to classify their severity, to check how long it takes the social media and/or big tech firm to take the violation down, and to monitor the overall compliance of the firm. Without such digital tools, the current situation will be perpetuated in which, after an algorithmic pre-screening, a human content moderator has to check content manually[45] and effective enforcement by regulators will prove nearly impossible.

IV. Integrating data science into regulatory agencies

Several competition authorities and other regulators across the globe have invested in integrating data science into their monitoring and

[42]Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) [2023] OJ L 2023/2854 and Commission Staff Working Document on Common European Data Spaces, 24 January 2024, SWD(2024) 21 final.

[43]Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) OJ [2024] L 2024/1689.

[44]The AI Act works with four levels of risk: unacceptable risk, high risk, limited risk, and minimal (or no) risk.

[45]Note that this comes at high mental and health costs as the job of content moderators has one of the highest occurrences of mental illness and trauma. See Miriah Steiger and others, 'The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support' (2021) Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems 1 <https://crowd.cs.vt.edu/wp-content/uploads/2021/02/CHI21_final__The_Psychological_Well_Being_of_Content_Moderators-2.pdf>.

enforcement activities. To support the enforcement of the DSA, the European Commission has set up the European Centre for Algorithmic Transparency (ECAT) hosted by the Joint Research Centre in close cooperation with DG Connect.[46] The UK Competition & Markets Authority (CMA) is probably the most far advanced with its Data, Technology and Analytics (DaTA) unit that was announced in October 2018.[47] Another worthwhile example is the *Pôle d'expertise de la régulation numérique* (PEReN) set up by the French government in August 2020. The distinctive feature of PEReN is that it is a team of experts not embedded in a particular agency but available to regulators and ministries in France upon request.[48] This section will discuss different ways of integrating data science into regulatory agencies and reflect on some best practices that can be identified so far.

A. Setting up data science teams

To build up data science expertise, a first step can be for regulators to start using data science tools internally to analyse own output and guidelines in order to structure internal processes and decision-making. This could for instance include a network analysis of own decisions to detect patterns, for instance with the aim of identifying similarities and differences in decisions targeting the same behaviour or the same industry. A next step is to move from this inward-directed use of data science to an outward-directed dimension by employing data science tools for monitoring and enforcement. Hunt distinguishes between four types of skills held by technologists of relevance to regulators, namely data science, data engineering, technology insight and behavioural insight.[49]

Data science skills allow for the extraction of insights from data using machine learning or AI techniques. These skills can be used in individual cases or for building tools for more advanced data techniques. Data

[46]See <https://algorithmic-transparency.ec.europa.eu/index_en>.

[47]Stefan Hunt, 'CMA's New DaTA Unit: Exciting Opportunities for Data Scientists' (24 October 2018) <https://competitionandmarkets.blog.gov.uk/2018/10/24/cmas-new-data-unit-exciting-opportunities-for-data-scientists/>. Other regulators across the European Union have hired data scientists as well, for instance the French *Autorité de la concurrence* and the Netherlands Authority for Consumers & Markets: see <https://www.autoritedelaconcurrence.fr/en/communiques-de-presse/autorite-creates-digital-economy-unit> and <https://www.acm.nl/nl/organisatie/werken-bij/vakgebieden/data-science>.

[48]See <https://www.peren.gouv.fr/en/qui-sommes-nous/>.

[49]Stefan Hunt, 'The Technology-Led Transformation of Competition and Consumer Agencies: The Competition and Markets Authority's Experience' (2022) UK Competition & Markets Authority Discussion Paper <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1085931/The_technology_led_transformation_of_competition_and_consumer_agencies.pdf>.

engineers focus on handling infrastructure and data pipelines and work for instance on security or the formatting of data. Both of these skills point at quantitative work in the form of coding and analysis of data. The other two types of skills relate to more qualitative work. Technology insight involves the analysis of technical issues relevant to the monitoring of markets or a specific case to support case teams or policy offers in understanding the implications of a particular technology. Technology insight is said to offer the highest immediate return in resources, because a small investment can benefit the understanding of an issue significantly.[50] Finally, behavioural insight develops findings about the behaviour of people and their decision-making from a range of disciplines such as behavioural economics, psychology, neuroscience and ethnography.[51]

Different choices can be made regarding how to set up a data science team. Most regulatory agencies that have started to invest in data science hired data scientists or technologists themselves and integrated their roles into their own agency. This is also the model used by the UK CMA for its DaTA unit. The French government took a different approach by establishing PEReN as an independent team of experts available to all regulatory agencies and ministries. The advantage of this approach is that the expertise is concentrated within a central entity, which may allow for more effective use of resources and joint development of state-of-the-art technological insights. This may especially be useful in the first stages of developing data science expertise in a particular jurisdiction. Because a regulatory agency is bound by its legal competences, fragmentation in technical knowledge may occur when its data scientists or technologists focus only on a sub-set of concerns (for instance competition or data protection concerns) or markets (for instance telecom or energy markets) depending on the mandate of the agency. A central entity like PEReN does not have such limitations and may therefore develop a more comprehensive range of data science expertise and technology insight. However, once experience grows, there may be a need to develop specific data science expertise and embedding data scientists in particular regulatory agencies may become more desirable. A centralized approach towards setting up data science expertise following the example of PEReN may thus be worthwhile to consider for jurisdictions in the initial stages.

[50]ibid 39.
[51]ibid 33–35.

B. How to involve data scientists into monitoring and enforcement work

Data scientists and technologists can be involved in different stages of monitoring and enforcement work, including at the stage of investigating an infringement, the design of remedies or the general monitoring of compliance with the law. For the design of remedies and general monitoring tasks, technology and behavioural insight can be particularly useful in order to assess whether suggested remedies can indeed address the identified concerns and to detect possible infringements. With regard to involvement in specific cases and investigations, the use of data science insights can lead to a better and more comprehensive understanding of internal documents and information provided by market players, especially on technical matters. Data scientists can for instance work with a case team to draft requests for information and to sit in meetings with parties to immediately verify whether technical statements are correct.[52] While the integration of data scientists into case teams allows them to become fully acquainted with the details of an investigation, it may sometimes be more effective to embed team members into a case only at particular stages upon need. Because the work on a case goes beyond what is relevant from a data science perspective, it can be both more efficient resource-wise and more interesting for data scientists or technologists to collaborate in a central unit or hub. The latter issue of motivation is an important aspect in order to retain data scientists for monitoring and enforcement work at regulatory agencies.[53] Both the UK CMA DaTA unit and PEReN rely on this model where data scientists are part of a central unit or hub and collaborate with case teams on particular investigations or issues where relevant.

This model also has similarities with how the Chief Economist Team is set up at the European Commission's Directorate General (DG) for Competition. The Chief Economist Team was established in 2003 to strengthen the economic underpinnings of competition analysis. The Chief Economist Team was envisaged to have two roles: (1) a support role by providing case teams with economic guidance and methodological insights, and (2) a checks-and-balances role by providing the Commissioner for Competition with an independent opinion before the adoption of a decision by

[52]ibid 16.
[53]ibid 40.

the College of Commissioners.[54] Considering that the integration of data science is a next multidisciplinary frontier in enforcement at competition and other regulatory agencies, inspiration may be drawn from the experience with the Chief Economist Team at the Commission's DG Competition. One best practice in this regard is to ensure that the involvement of data scientists is based on a clear mandate and organizational structure because of the difference in vocabulary and methodologies. For competition authorities, this may be more intuitive. Not only because competition authorities already have experience with integrating economics as a quantitative form of input into their analysis, but also because the techniques used by data scientists often build on econometrics and can sometimes substitute the use of economic tools.[55]

C. Data science as a way to move from reactive to proactive enforcement

Beyond supporting regular monitoring and enforcement tasks, data science has the potential to truly reform the work of regulatory agencies. While enforcement is now mostly reactive with regulators acting upon complaints or own targeted investigations, systematic data gathering through the use of data science tools can lead to a more proactive form of enforcement.[56] This would allow for earlier identification of harm, so that interventions to restore competitive conditions or address consumer detriment can be more effective and less costly.[57] In addition, proactive enforcement can lead to increased deterrence because of the structured and comprehensive monitoring of behaviour that no longer depends on signals from market players or consumers.[58]

To engage in more systematic monitoring, regulatory agencies will need to develop own tools and technologies because commercially available options typically do not fully meet their needs.[59] Because of the overlap across activities and competences, it is desirable for regulators to share data science expertise and tools within and across jurisdictions.

[54] Lars-Hendrik Röller and Pierre A Buigues, 'The Office of the Chief Competition Economist at the European Commission' (2005) 6–9 <https://competition-policy.ec.europa.eu/document/download/284bc523-763d-4717-848d-e1a1e636ae86_en?filename=officechiefecon_ec.pdf>.
[55] Hunt (n 49) 34–36.
[56] See also Christina Riefa, 'Transforming Consumer Law Enforcement with Technology: From Reactive to Proactive?' (2023) 12(3) Journal of European Consumer and Market Law 97.
[57] Hunt (n 49) 30.
[58] ibid 46.
[59] ibid 23. See also Christina Riefa, 'Transforming Consumer Law Enforcement with Technology: From Reactive to Proactive?' (2023) 12(3) Journal of European Consumer and Market Law 97, 100.

D. Desirability of sharing expertise and tools across jurisdictions

Although regulators may to some extent have own demands for the specific data science expertise and tools they need in their domain, they will often monitor similar behaviour and problems. Because the development of own technologies requires large investments that only pay off in the medium or long term, the potential gains of collaboration are substantial.[60] Joining forces can happen at a global scale, because monitoring needs will largely overlap even though the applicable legal frameworks differ. With its role in the enforcement of the DSA and the DMA, the European Commission could take the lead in ensuring effective exchange of data science expertise and tools across Member States.

Coding is an example of an area where exchange of tools seems particularly promising. If regulators would share code with each other for scraping or analysis where the transfer of code and know-how has low costs, the returns on investment in developing these tools can multiply. By making tools available to each other, regulators can jointly become more effective. Sharing of codes could even happen in open source to allow for possible technical errors to be more quickly detected and remedied. Such forms of exchange and collaboration among regulators also enable agencies with less resources or less ability to invest in the development of technology to benefit from data science expertise and tools. The potential of international collaboration among regulators in the area of data science therefore carries large potential and deserves to be further explored.[61]

V. Conclusion

A range of data science tools is already available and used by competition authorities and other regulators across Europe and beyond. These tools are also of relevance for checking compliance with the DSA and DMA. Data science tools can help to overcome challenges in terms of data collection, data processing and analysis, and data visualization for regulatory agencies and other third parties wishing to study the behaviour of platforms. However, data science tools also have limitations in that they do not provide a complete picture but rather a sample of insights and that absolute conclusions cannot be drawn due to the probabilistic nature of most techniques.

[60]Hunt (n 49) 26.
[61]ibid 45–46.

The growing use of data science tools by competition authorities and other regulators holds promise to make enforcement more effective, but also comes with risks. Data science has the potential of leading to the democratization of investigations by allowing regulatory agencies with less resources to do similar investigations as better endowed agencies. At the same time, the increasing collection and use of data requires attention for the protection of privacy and the transparency and explainability of decisions.

So far, the most effective way of integrating data scientists into the organizational structure of regulators seems to be in the form of a central hub either at the level of particular agency (like the UK CMA DaTA Unit) or at the level of a country (PEReN). Data science can be integrated into individual cases to provide a better understanding of the impact of the behaviour in question, but it also carries the potential of reforming the work of regulators by moving towards more systematic monitoring of markets going beyond mere reactive enforcement. To achieve this, the exchange of data science expertise and tools among regulators deserves to be further facilitated within and across jurisdictions.

Acknowledgements

We are especially grateful for the excellent input delivered by the contractor, Visionary Analytics, to the work of the expert group for the EU Observatory on the Online Platform Economy, on which this paper is based. In addition, we would like to thank Laura Edelson, Stefan Hunt, Damien Neven, and Patricia Prüfer for participating in a closed workshop at the European Commission on 25 April 2023, and Asia Biega, Nicolas Deffieux, Louis-Victor de Franssu, and Catalina Goanta for participating in a public workshop on 27 June 2023 during the European Commission's DSA Stakeholder Event.

Disclosure statement

This paper is based on work that the authors have done as part of the expert group for the EU Observatory on the Online Platform Economy and builds on the input delivered by the contractor, Visionary Analytics. All contents, opinions and errors are our own and do not reflect the position of the European Commission.

ORCID


Inge Graef http://orcid.org/0000-0001-5543-615X