
UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI ECONOMIA "GIORGIO FUÀ"

Corso di Laurea triennale in

**Digital Economics and Business**

# Personal Data as an Economic Asset in the Digital Economy: Legal Limits in the Context of AI and Online Platforms.

Relatore:

Prof. Roberto Ruoppo

Rapporto Finale di:

Rebecca Lasca

Anno Accademico 2025/2026

## TABLE OF CONTENTS

# INTRODUCTION

The development of the digital economy has profoundly changed the role of personal data. In the past, personal information was collected mainly in limited contexts, such as banking, healthcare or institutional relationships. Today, however, data are continuously generated through everyday interactions with digital platforms, online services and artificial intelligence systems. Search engines, social media platforms, mobile applications and recommender systems collect and process large amounts of information concerning users' behaviors, preferences, interests and interactions. This thesis analyses personal data both as an economic resource and as an object of legal protection. On the one hand, personal data have become a strategic asset for digital platforms, since they support targeted advertising, personalization, profiling and the development of AI-based services. On the other hand, their collection and exploitation raise important concerns regarding privacy, transparency, individual autonomy and fundamental rights. The first chapter examines the economic value of personal data, focusing on the process through which data are transformed into assets and on the difficulties connected with their monetary valuation and taxation. The second chapter analyses the collection and use of personal data by digital platforms and AI systems, with particular attention to profiling and recommender systems. The third chapter focuses on the legal limits imposed by the European regulatory framework, especially the GDPR, the Digital Services Act, the Digital Markets Act and the Artificial Intelligence Act. Finally, the fourth chapter examines the Clearview AI

case, which represents a significant example of the legal risks connected with the large-scale exploitation of biometric data. The aim of the thesis is to show that personal data cannot be considered only as an economic commodity. Although they generate significant value in digital markets, they remain closely connected to individuals' identity, autonomy and fundamental rights. For this reason, their economic use must be balanced with effective legal safeguards.

# PERSONAL DATA AS AN ECONOMIC ASSET

## 1.1 ASSETIZATION OF PERSONAL DATA

The contemporary digital economy is characterized by an unprecedented expansion in the collection and use of personal data. Digital platforms, artificial intelligence systems, and online services continuously generate and process vast amounts of information relating to individuals' behaviors, preferences, and interactions. In recent years, personal data have often been described as the new 'oil' of the digital economy, reflecting their strategic importance for innovation, competition, and value creation in digital markets. While this metaphor has been widely debated, it highlights the growing recognition that data constitute a fundamental input for the development of new business models and technological advancements. Personal data are increasingly framed as a distinct economic asset for large technology firms and investors. The key issue is not simply the existence of personal information, but how it is transformed into an economically exploitable resource capable of generating revenue. Traditionally, personal data such as names, addresses, or social relationships have been regarded as information rather than creative outputs. Consequently, they have not been subject to ownership in the same way as intellectual property or tangible goods. Nevertheless, the absence of formal ownership does not prevent personal data from acquiring economic significance. The concept of assetization refers precisely to the transformation of data into capitalized resources through organizational, technological, and strategic practices that enable firms to structure, measure, and monetize access to user

information over time. In this context, the economic value of personal data does not arise from their intrinsic characteristics, but from the ability of digital firms to enclose data within controlled ecosystems and to regulate their access and use. Before becoming economically relevant, data must be collected, standardized, and analyzed in ways that make them measurable and operational within market-oriented business strategies. In this context, personal data are increasingly treated as tradable economic goods within digital markets. However, in most cases, data are not directly sold by individuals, but rather by organizations that collect and exploit user information for commercial purposes, often through contractual agreements embedded in digital services. [1] Digital platforms monetize user participation by transforming continuous flows of behavioral data into stable revenue streams, linking the value of personal data to firms' ability to maintain long-term user relationships and systematically extract insights that support profitability and market expansion.[2] Moreover, the economic value of personal data is closely linked to the control exercised by digital platforms over their access and use. In digital markets, certain large platforms are identified as 'gatekeepers', due to their significant impact on the internal market and their central position in connecting users and businesses. These platforms operate as key intermediaries, controlling access to core platform services and the data generated within their

1 BIRCH K., COCHRANE D., WARD C., Data as asset? The measurement, governance, and valuation of digital personal data, 'Big Data & Society', 2021.
2 BATAINEH A. S., MIZOUNI R., EL BARACHI M., BENTAHAR J., Monetizing Personal Data: A Two-Sided Market Approach, 'Procedia Computer Science', Vol. 83, 2016, pp. 472– 479.

ecosystems. As recognized in the Digital Markets Act, such control over data contributes to strengthening their market position and reinforces their ability to shape competitive dynamics. In particular, the concentration of data within these platforms limits the ability of other market actors to access and use comparable information, thereby consolidating the economic and strategic value of personal data as a controlled resource rather than an openly accessible one. Before analyzing the economic dynamics of data assetization, it is useful to briefly frame the legal definition and protection of personal data within the European context.[3] Personal data are defined in Article 4(1) of the General Data Protection Regulation as 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'. This broad definition reflects the increasing relevance of personal data in the digital environment, where even indirect identifiers may enable the identification of individuals.[4] The protection of personal data is closely related to, but distinct from, the right to privacy. While the latter traditionally concerns the protection of an individual's private sphere against external interference, the former has developed as a specific legal framework governing the collection and

3 BIRCH K., COCHRANE D., WARD C., Data as asset? The measurement, governance, and valuation of digital personal data, 'Big Data & Society', 2021.
4 EUROPEAN PARLIAMENT AND COUNCIL, Regulation (EU) 2022/1925 (Digital Markets Act), 2022, Art. 3 and Art. 6.

use of personal information. This distinction became particularly relevant with the rise of the information society and the growing capacity to process large amounts of data as reflected in Articles 7 and 8 of the Charter of Fundamental Rights of the European Union. In the European context, personal data protection is explicitly recognized as a fundamental right by Article 8 of the Charter of Fundamental Rights of the European Union, which establishes that personal data must be processed fairly, for specified purposes, and on a legitimate basis, under the supervision of an independent authority. This provision constitutes the legal foundation of modern data protection law. Within this framework, the processing of personal data must comply with specific legal conditions.[5] Consent represents one of the primary legal bases for data processing under Article 6(1)(a) of the GDPR.[6] According to Article 4(11), consent must be freely given, specific, informed, and unambiguous. These requirements ensure that individuals retain effective control over their personal data, including the possibility to grant, refuse, or withdraw consent at any time.[7]In the context of digital platforms, however, the practical implementation of consent has raised significant concerns. The practical application of these principles has been further clarified by the European Data Protection Board[8], whose guidelines specify that valid consent requires a clear affirmative action and can be withdrawn at any time. These interpretative tools play a crucial role in ensuring a consistent application of data protection rules

5 Charter of Fundamental Rights of the European Union, 2012/C 326/02, Arts. 7 and 8.
6 Regulation (EU) 2016/679 (General Data Protection Regulation), Art. 6(1)(a).
7 Regulation (EU) 2016/679 (General Data Protection Regulation), Art. 4(11).
8 EUROPEAN DATA PROTECTION BOARD, Guidelines 05/2020 on consent under Regulation 2016/679, 2020.

across Member States. [9]Over time, the way personal data is perceived and managed has undergone a significant transformation. In the past, personal data were shared only in specific and limited situations, such as interactions with institutions like banks or healthcare providers. In these contexts, data collection was occasional and circumscribed, and individuals maintained a relatively high degree of control over when and how their information was disclosed. As a result, personal data were less widespread, less accessible, and closely linked to moments of social and economic life. In contrast, in the contemporary digital environment, personal data is continuously generated and collected through everyday activities. Modern technologies enable constant recording and storage of information, making it increasingly difficult for individuals to avoid producing digital traces. Consequently, it has become almost impossible to exist without leaving some form of data record associated with one's actions. Personal data are no longer confined to specific situations, but have become pervasive, continuously produced, and often collected independently of the individual's full awareness. This transformation highlights a shift from a context in which personal data were limited, controllable, and situational, to one in which they are widespread, continuously generated, and increasingly difficult to control. This shift highlights a transition from a model based on limited and controlled data sharing to one characterized by continuous data generation and reduced individual control.[10] The process through

10 CIFALDI G., Evolution of Privacy and Data Protection in the Digital Era, 2023

which personal data are transformed into economic assets is closely connected to the emergence of what Zuboff defines as surveillance capitalism. Within this economic model, digital firms systematically extract behavioral data from users' interactions, converting human experience into a new form of raw material for economic production. This behavioral information, often generated as a secondary by-product of online activities, is analyzed and processed to develop predictive products capable of anticipating future actions. Consequently, the monetization of personal data relies on the creation of new digital markets in which behavioral predictions are traded as valuable economic resources. In the early stages of digital platform development, companies such as Google collected behavioral data primarily to improve the performance and relevance of their services. These data traces, frequently described as 'data exhaust', were initially considered operational by-products rather than strategic assets. Over time, however, firms began to recognize their economic potential. The systematic capture and analysis of behavioral signals enabled the creation of recursive feedback loops, through which user interactions contributed to continuous technological learning and algorithmic optimization. At this initial stage, the use of behavioral data was largely oriented towards enhancing user experience. Zuboff characterizes this dynamic as a 'behavioral value-reinvestment cycle', in which data generated by users were reinvested in service improvements rather than directly monetized. Platforms and users were thus engaged in a symbiotic relationship: firms relied on behavioral data to refine their technologies, while users benefited from more accurate and

accessible digital services. However, this phase also laid the foundations for subsequent transformations in platform business models, as firms gradually realized that behavioral data could be repurposed beyond service optimization. Over time, financial pressures and competitive dynamics encouraged digital firms to adopt a more explicit commercial orientation.

Behavioral data that exceeded what was necessary for improving services began to be reinterpreted as behavioral surplus, a residual category of information that could be exploited for profit generation. By analyzing patterns of online activity, platforms were able to infer users' preferences and intentions with increasing precision, constructing detailed user profiles that enabled highly personalized advertising strategies. Behavioral surplus thus became a strategic economic resource, supporting new forms of data-driven accumulation and encouraging firms to expand the scale and scope of data extraction across multiple domains of users' lives. Ultimately, the assetization of personal data reflects a broader transformation in contemporary capitalism, in which economic value increasingly derives from the continuous extraction, analysis, and monetization of behavioral information. Understanding these processes is essential for examining both the strategic role of data in digital markets and the emerging challenges associated with data-driven platform business models. This perspective leads to the

examination of how digital platforms construct data as calculable resources through specific systems of measurement and quantification. [11]

### 1.2 VALUE OF PERSONAL DATA AND TAXATION CHALLENGES

In the digital economy, data represents a central component of the value creation process. As highlighted by the OECD, information collected from users constitutes a key input in the development of digital products and services, enabling firms not only to improve their offerings, but also to refine their ability to segment consumers and optimize decision-making processes. At the same time, users themselves actively contribute to value creation, for instance through content generation, interaction with digital platforms and the continuous production of behavioral data, thereby increasing the utility and attractiveness of the services provided. In this sense, both data and user participation become essential elements within the value chain of digital firms, contributing to the generation of economic value in ways that are often indirect and difficult to quantify. However, the emergence of these new forms of value creation raises significant challenges for traditional tax systems. It becomes increasingly difficult to determine where economic value is generated and, consequently, which jurisdiction has the right to tax it. Existing tax rules are largely based on the physical presence of firms within a given territory, whereas digital business models allow companies to operate simultaneously in multiple markets without a physical material presence, making the allocation of taxing rights across countries more uncertain. A further issue concerns the difficulty of assigning an economic value to data themselves. Although firms can collect, analyze and monetize large volumes of data, it remains complex to determine an objective value of raw data for

11 ZUBOFF S., The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power, PublicAffairs, New York, 2019.

tax purposes, especially because such data are generally not recognized as assets in financial statements and their value depends significantly on the ways in which they are processed, combined and used within specific business models. This uncertainty complicates both the measurement of value creation, and it's appropriate taxations tax authorities lack clear benchmarks for assessing the economic contribution of data-driven activities. Finally, the digital economy also raises important challenges in relation to indirect taxation. In particular, the collection of value added tax (VAT) on digital transactions presents significant difficulties, not only in terms of revenue collection, but also in the effective application and enforcement of tax rules. These challenges are especially evident in the case of digital services and cross border transactions, where the absence of a physical presence and the intangible nature of economic activities make monitoring, compliance, and enforcement more complex. In this context, the so-called VAT gap, namely the difference between expected and collected tax revenues, further highlights the limitations of current tax systems in capturing the value generated within the digital economy.[12] From a VAT perspective, a central legal issue concerns the concept of consideration, which traditionally requires a direct monetary payment for a transaction to be taxable. However, this framework becomes problematic in the context of digital services, where transactions often do not involve any explicit financial payment. In many cases, digital platforms provide services that appear to be free of charge. Nevertheless, these services are not entirely without cost, as users grant access to their personal data in exchange for the use of such services. As highlighted in the literature, digital business models are often based on so called 'zero-price' services, where users effectively pay with

12 OECD, Addressing the Tax Challenges of the Digital Economy, Action 1 – 2015 Final Report, OECD Publishing, Paris, 2015

their personal data rather than money, even if such payment is not immediately visible or explicit recognized as such. This raises an important legal question as to whether personal data can be considered a form of consideration, or counter-performance, within the meaning of VAT law. If personal data are interpreted as a form of payment, the exchange between users and digital platforms could potentially fall within the scope of taxable transactions. At the same time, the literature also underlines that personal data cannot be treated as a simple commodity, given their connection to fundamental rights and data protection principles within the European legal framework. The absence of a clear monetary payment therefore creates uncertainty in the legal qualification of these exchanges, as traditional VAT rules are not designed to capture non-monetary forms of value transfer. As a result, the distinction between free services and taxable transactions becomes increasingly blurred, highlighting the limits of existing legal frameworks in addressing data driven business models.[13] A concrete example of these issues can be observed in the Italian context, where tax authorities have recently focused on the VAT treatment of digital platforms. Investigations have targeted major technology companies such as Meta and X, questioning whether services commonly considered 'free' should instead be treated as taxable transactions. According to the Italian Tax Authority, when users register on a platform without monetary payment but provide personal data, such interaction may constitute a form of barter transaction, in which data represent a non-monetary consideration. On this basis, the provision of digital services could fall within the scope of value added tax (VAT), provided that a direct link between the service and the consideration can be established. However, this interpretation remains controversial

13 MURSIA M., TROVATO C. A., The commodification of our digital identity: limits on monetising personal data in the European context, 'Rivista di Diritto dei Media', 2021.

at the European level. The VAT Committee has previously clarified that such transactions may not be taxable where no direct and identifiable link can be established between the service provided and the personal data received. In particular, the difficulty of determining the value of personal data and the absence of a clear contractual exchange raise significant doubts about the applicability of VAT in these cases. Overall, the Italian case highlights the broader challenges faced by tax systems in adapting existing VAT rules to digital business models, as well as the risk that significant economic value generated through user data may not be effectively captured within current tax frameworks.[14]

### 1.3 METHODOLOGIES FOR ESTIMATING THE MONETARY VALUE OF PERSONAL DATA

Estimating the monetary value of personal data represents a complex analytical challenge, as there is no universally accepted methodology capable of capturing the economic worth of data across different contexts. Existing approaches primarily focus on identifying the prices that markets assign to personal data rather than measuring their broader social or economic benefits. In practice, most valuation methods attempt to determine the value associated with an individual data record or user profile, although the definition of what constitutes a 'record' may vary significantly depending on the business model and the type of information involved. A data record can refer to a single piece of personal information, such as age or location, or to a comprehensive profile including demographic, financial and behavioral characteristics. Consequently, the estimated value of personal data is inherently context-dependent and cannot be generalized across markets or industries. Current methodologies can be broadly divided into two main categories:

14 GIULIANO S., FEDERICI M., Italy's Tax Authorities Target 'Free' Online Services for VAT, Bloomberg Tax, 2025.

approaches based on market valuation and approaches based on individuals' perceptions of value. Market-based methods rely on observable economic indicators derived from transactions or financial performance, while individual-based methods focus on how users subjectively evaluate the economic importance of their personal information and privacy. Although these approaches provide useful preliminary insights, they measure different dimensions of value, making direct comparisons difficult. Furthermore, personal data cannot be analyzed in isolation from the specific digital business models in which they are generated and exploited, as their economic significance depends heavily on contextual factors such as regulatory frameworks, technological infrastructures and patterns of data use. Market-based methodologies attempt to derive an approximate monetary value of personal data from observable financial and transactional indicators. One frequently proposed approach consists in calculating financial results per data record, obtained by dividing a company's market capitalization, revenues or net income by the number of personal data records it exploits. This method provides a preliminary proxy of the economic contribution generated by personal data. Another valuation technique relies on market prices for personal data observed in transactions involving data brokers. In this case, the value of a specific data entry is inferred from its price in exchange contexts, even though such prices often include additional costs related to data collection, search and processing, and may vary considerably according to demand conditions and intended use. Further estimates can be derived from analyzing the economic costs associated with data breaches, which reflect the financial losses incurred by firms or individuals following the unauthorized disclosure of personal information. While this indicator highlights the risks that organizations must mitigate, it captures the cost of potential damage rather than the intrinsic value of the data themselves. Finally, some estimates of the value of personal

data are based on prices observed in illegal markets, where stolen information is traded. Although these values may provide an indication of the practical use of personal data, they are not fully reliable, as they are difficult to measure and may underestimate the real value due to the risks associated with illegal activities. In addition to market-based approaches, other methodologies focus on how individuals themselves perceive the value of their personal data. These methods typically rely on surveys and economic experiments, in which users are asked to assign a monetary value to specific types of personal information or to evaluate trade-offs between privacy protection and access to digital services. Through these tools, it is possible to understand how the perceived value of personal data varies depending on the context and the type of information considered. A related indicator is individuals' willingness to pay to protect their personal data. This approach measures how much users would be willing to spend to prevent unauthorized access or misuse of their information, providing further insight into the subjective value attributed to privacy. Overall, these methodologies show that the value of personal data cannot be considered fixed or intrinsic.[15] This uncertainty in valuation is not only an economic issue but also has relevant implications in legal contexts where a precise monetary assessment is required. Under Article 82 of the GDPR, individuals have the right to compensation for both material and non-material damage resulting from data processing. However, the absence of a clear and consistent monetary valuation of personal data makes it difficult to quantify such damages in practice.[16]

15 OECD, Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value, OECD Digital Economy Papers, n. 220, OECD Publishing, Paris, 2013.
16 EUROPEAN PARLIAMENT AND COUNCIL, Regulation (EU) 2016/679 (General Data Protection Regulation), 2016, Art. 82.

# COLLECTION AND USE OF PERSONAL DATA IN DIGITAL PLATFORMS AND AI SYSTEMS

## 2.1 DIGITAL PLATFORMS AND THE ROLE OF DATA IN DIGITAL ECOSYSTEMS

The analysis conducted in the previous chapter has shown how personal data have progressively evolved into a key economic asset within digital markets. However, their value does not stem solely from their existence or potential exchange value, but rather from the concrete mechanisms through which they are collected, processed, and exploited within digital ecosystems. To understand the regulatory framework governing these processes, it is necessary to clarify the main categories of intermediary services defined under Article 3 of the Digital Services Act. In this regard, a 'service intermediary' refers to one of the information society services falling within the scope of intermediary services, including online platforms and online search engines. Within this framework, an 'online platform' is defined a 'a hosting service which, at the request of a recipient of the service, stores and disseminates information to the public, unless such activity is a minor and purely ancillary feature of another service or a minor function of the principal service, and, for objective and technical reasons, cannot be used without that other service'. Another category defined by the Digital Services Act is that of 'online search engine', which is defined as 'an intermediary service that allows users to input queries in order to perform searches, in principle, across all websites or across websites in a particular language, based on a query expressed in the form of keywords, voice commands, phrases, or other input, and which returns results in any format in which information related to the requested content can be found.' [17] These services have become among the most

17 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act), amending Directive 2000/31/EC, Article 3(i), OJ L 277, 27.10.2022.

significant global processors of personal data, due to their central role in contemporary digital markets and their pervasive integration into everyday digital interactions. This development has intensified legal and regulatory concerns regarding the appropriate balance between the protection of individuals' personal data and the operational freedom of digital service providers and other users. In response to these challenges, the European Union has developed a comprehensive regulatory framework, of which the Digital Services Act forms a central component. The DSA is part of the broader European Digital Strategy, which also includes instruments such as the General Data Protection Regulation (GDPR), the Digital Markets Act (DMA), the Artificial Intelligence Act. Building on the foundations of the e-Commerce Directive of 2000, the DSA further refines the legal categorization of intermediary services and updates the regulatory approach to digital platforms.[18] Against this background, it is necessary to examine how these actors concretely collect and process user data in practice. The collection of user data is primarily motivated by two interconnected objectives: the improvement of service quality and the monetization of digital services, particularly through targeted advertising. The types of data collected are highly heterogeneous. Online providers typically gather technological data, such as IP addresses and device identifiers, which are automatically generated during user interaction with digital services. In addition, they collect personal data, including demographic information and user preferences, as well as behavioral data derived from online activity, such as browsing patterns, search queries, and purchasing behavior. A significant part of this data collection process relies on tracking technologies such as cookies, which function as digital identifiers enabling the monitoring of user

18 FARINHO D. S., Personal Data Processing by Online Platforms and Search Engines: The Case of the EU Digital Services Act, 'Public Governance, Administration and Finances Law Review', Vol. 9, No. 1, 2024, pp. 37–58

activity across platforms and websites. Collected data are then used for multiple operational purposes. In the context of search engines and other search-based services, user-generated data such as queries and click behavior are analyzed to improve the relevance and quality of search results. These signals allow providers to identify patterns of user preferences and refine ranking mechanisms accordingly. Moreover, online providers frequently test modifications to interfaces and service features, using behavioral data to evaluate user interaction and assess the effectiveness of such changes. Beyond service optimization, user data plays a central role in personalization processes and recommendation systems by analyzing behavioral and contextual information, online providers can suggest relevant content, products, or services, thereby enhancing user engagement and platform efficiency. This capacity to personalize services is closely linked to the development of data-driven advertising models, where the economic value of platforms depends on their ability to deliver targeted advertisements. In this sense, user data not only improves service quality but also strengthens the competitive position of online providers by increasing advertising revenues and supporting the provision of low-cost or free services to users.[19] In this context, the role of large digital platforms becomes even more significant, particularly regarding the so-called 'gatekeepers', which occupy a central position within digital ecosystems. Under Article 3(1) of the Digital Markets Act, an undertaking is designated as a gatekeeper when it satisfies a set of qualitative criteria. In particular, the undertaking must have a significant impact on the internal market, provide a core platform service that functions as an important gateway for business users to reach end users, and enjoy, or be expected to enjoy soon, an entrenched and durable

19 LERNER A. V., The Role of 'Big Data' in Online Platform Competition, 'SSRN Electronic Journal', 2014, pp. 7-20.

position in its operations.[20] These criteria suggest that gatekeepers occupy a structurally significant position within digital ecosystems, as they can intermediate between different groups of users and influence access to digital markets. As highlighted in the economic literature on digital platforms, exclusive access to large volumes of individual-level data may provide a significant competitive advantage, even when such data are used in anonymized form for algorithmic purposes. In particular, the availability of both historical and continuously generated data can improve the quality of services and enhance algorithmic performance. This dynamic may give rise to self-reinforcing feedback loops, whereby a superior service attracts more users, who in turn generate additional and more up-to-date data, further strengthening the platform's position. Moreover, platforms may expand their data collection capabilities through various mechanisms within digital ecosystems. For instance, data interoperability arrangements may require interconnected firms to share user data, allowing platforms to collect information beyond their own services and centralize it. Similarly, the use of internal application programming interfaces between services within the same ecosystem may create strong advantages for platforms operating across multiple and diverse services. At the same time, large platforms often act as intermediaries enabling third-party providers to offer complementary services on top of their infrastructure, thereby further expanding the scope of data collection and integration.[21]

20 Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector (Digital Markets Act), Article 3(1).
21 Jacques Crémer, Yves-Alexandre de Montoya and Heike Schweitzer, Competition Policy for the Digital Era, Report for the European Commission, Directorate-General for Competition, Publications Office of the European Union, Luxembourg, 2019.

## 2.2 PROFILING, DATA DRIVEN-PRACTICES AND LEGAL SAFEGUARDS UNDER THE GDPR

Building on the data collection and accumulation practices described above, digital platforms engage in more advanced forms of data processing, among which profiling plays a central role. Under Article 4(4) of the General Data Protection Regulation, 'profiling' refers to 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements'.[22] More broadly, profiling can be understood as the automated processing of data aimed at deriving, inferring, predicting or evaluating information about individuals, including their behavior, preferences, or characteristics. Profiling practices rely on the aggregation and analysis of large volumes of data collected from a variety of sources. Through these processes, platforms can generate new knowledge about individuals that goes beyond the information provided by users. Seemingly non-sensitive data, such as browsing behavior, location data, or interaction patterns, can be used to infer highly sensitive information, including aspects related to health, political opinions, personality traits, or socio-economic status. These inferences are made possible through increasingly sophisticated data processing techniques, including statistical analysis and machine learning algorithms, which enable the identification of patterns and correlations within large datasets. As a result, profiling allows platforms to construct detailed and comprehensive user profiles, which may include both observed and inferred

22 Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), Article 4(4).

characteristics. These profiles are not static, but are continuously updated as new data is collected, allowing platforms to refine their predictions over time. Profiling can be conducted at both the individual and group level and may rely either on data directly associated with a specific individual or on patterns derived from a broader population. In this sense, profiling enables the categorization and classification of individuals, often based on probabilistic assessments rather than verified information. The knowledge generated through profiling is then used to make or inform a wide range of decisions. In the context of digital platforms, this includes the personalization of content and services, recommendation systems, and targeted advertising, which shape users' online experiences by tailoring information to their presumed interests or preferences. More broadly, profiling can also support more consequential forms of decision-making, such as those related to employment, credit scoring, or access to services, sometimes with limited or no human intervention. In these cases, profiling may play a decisive role in determining opportunities and outcomes for individuals.[23] In this context, the use of personal data within digital platforms can be further understood as part of a structured process, often described as a three-stage cycle consisting of data collection, profiling, and microtargeting or personalization. The first stage, data collection, involves the continuous gathering of personal data from a wide variety of sources. While some data are actively provided by users, a significant portion is passively generated through everyday digital interactions, often referred to as 'digital breadcrumbs'. These include browsing behavior, location data, device identifiers, social media activity, and purchase history, which are collected through tracking technologies such as cookies, web beacons, and device

23 Frederike Kaltheuner and Elettra Bietti, Data is Power: Towards Additional Guidance on Profiling and Automated Decision-Making in the GDPR, (2018) 2(2) Journal of Information Rights, Policy and Practice.

fingerprinting. The increasing use of connected devices further expands the scope of data collection, allowing platforms to monitor users' behavior across different contexts and over time. The second stage, profiling, consists of the analysis and processing of collected data to segment individuals into increasingly detailed categories. Through statistical methods and computational techniques, platforms can infer preferences, behaviors, and personal characteristics, often going beyond the information directly provided by users. Profiling enables the identification of patterns that are not immediately visible and allows the construction of complex user profiles, which may include sensitive or inferred attributes. The final stage, micro-targeting or personalization, involves the use of these profiles to tailor content, messages, or services to specific individuals or groups. This may include personalized advertising, the ranking and selection of content in news feeds, or the customization of online experiences based on predicted preferences. In this phase, users' informational environments are shaped according to their inferred characteristics, often with the aim of maximizing engagement or economic returns. While such practices may vary in their impact, they can also raise concerns when they are used to influence behavior or decision-making in ways that are not transparent to the individual.[24] However, the increasing reliance on profiling raises significant concerns. The ability to infer sensitive information from large datasets challenges traditional notions of privacy and individual control over personal data. Moreover, profiling processes may lack transparency, making it difficult for individuals to understand how their data is used or how decisions affecting them are made. At the same time, the use of inferred or predicted data introduces risks of inaccuracy and bias, which may lead to individuals being

24 European Data Protection Supervisor (EDPS), Opinion 3/2018 on online manipulation and personal data, 19 March 2018.

misclassified, misjudged, or unfairly treated. When profiling is used to inform decisions that have significant effects on individuals, such risks may translate into concrete harms, including discrimination or unjustified limitations on access to services and opportunities. [25] In this context, the need for effective legal safeguards becomes particularly evident. The General Data Protection Regulation establishes a set of general provisions that apply to all forms of profiling and automated decision-making, through the data protection principles set out in Article 5 and the lawful bases for processing under Article 6. Among these, the principle of lawfulness, fairness and transparency, as provided under Article 5(1)(a), plays a central role. Profiling processes are often invisible to the data subject, as they rely on the creation of derived or inferred data that has not been directly provided by the individual. For this reason, controllers are required to provide clear, concise and easily accessible information about the processing of personal data in accordance with articles 12,13, and 14 of the GDPR, both at the time of data collection and, in the case of indirectly obtained data, within the time limits established by the Regulations. At the same time, processing must be fair, as profiling may lead to discriminatory outcomes, for instance by limiting access to employment, credit or other opportunities. [26] In addition, profiling must comply with the principle of purpose limitation under Article 5(1)(b), as personal data are often reused for purposes different from those for which they were originally collected. The compatibility of such further processing depends on several factors, including the information initially provided to the data subject. The principle of data

25 Frederike Kaltheuner and Elettra Bietti, Data is Power: Towards Additional Guidance on Profiling and Automated Decision-Making in the GDPR, (2018) 2(2) Journal of Information Rights, Policy and Practice.
26 Article 29 Data Protection Working Party, Guidelines on Automated Individual Decision-making and Profiling for the purposes of Regulation (EU) 2016/679, WP251rev.01, adopted on 3 October 2017 and last revised and adopted on 6 February 2018

minimization under Article 5(1)(c) further requires that controllers limit the collection and retention of personal data to what is strictly necessary, despite the incentives created by technological developments to collect and store large volumes of data. In this context, controllers should be able to justify the need for the data used in profiling or consider alternative approaches, such as anonymization or pseudonymization where appropriate. Accuracy, as required by Article 5(1)(d), also represents a crucial requirement, as profiling involves multiple stages, including data collection, analysis, profile creation, and the application of profiles in decision-making processes. Inaccurate data at any of these stages may result in flawed profiles and incorrect outcomes affecting individuals.[27] Beyond these principles, profiling must rely on a valid legal basis under Article 6 of the GDPR. Consent, as provided under Article 6(1)(a), also represents one possible basis, but its use is particularly complex in the context of profiling, as individuals must be able to fully understand the nature and implications of the processing, including the use of inferred data. Finally, profiling may involve the processing of special categories of data under Article 9, even when such data is not directly collected but inferred through correlations. In such cases, controllers must ensure that the processing is compatible with the original purpose, identify an appropriate legal basis, and provide adequate information to the data subject. The GDPR establishes a set of rights aimed at protecting data subjects in the context of profiling and automated decision-making. These rights apply

27 Article 29 Data Protection Working Party, Guidelines on Profiling, WP251rev.01, 2018.

broadly to profiling activities, regardless of whether they result in automated decisions, and are enforceable against both the controller responsible for creating the profile and the controller making decisions based on it. In particular, the right to be informed, as provided under Articles 13 and 14 of the GDPR, plays a central role in ensuring transparency. Controllers are required to clearly and simply explain how profiling and automated decision-making processes operate. Where profiling is used to support decision-making, individuals must be explicitly informed about both the existence of profiling and th fact that decisions may be taken based on the profiles generated.[28]In addition, data subjects benefit from the right to rectification under Article 16 and the right to erasure under Article 17. These rights apply not only to the personal data used as input for profiling, but also to the resulting output data, such as the profile itself or any score assigned to the individual. Article 16 also allows individuals to complement incomplete personal data with additional information. Furthermore, Article 21 provides individuals with the right to object to processing, including profiling, on grounds relating to their situation. Controllers are required to explicitly inform data subjects of this right and present it clearly and separately from other information. Finally, additional safeguards are provided under Article 22, which grants individuals the right not to be subject to decisions based solely on automated processing, including profiling, where such decisions produce legal effects or similarly significant consequences. However, this provision applies only in specific circumstances and

[28] Article 29 Data Protection Working Party, Guidelines on Profiling, WP251rev.01, 2018

does not cover all forms of profiling.[29] Overall, these rights aim to ensure that individuals retain a degree of control over their personal data and can challenge profiling practices within digital environments. However, the practical effectiveness of these safeguards may be limited by the complexity and opacity of profiling techniques, which can make it difficult for individuals to fully understand and exercise their rights in practice.

### 2.3 ARTIFICIAL INTELLIGENCE, PERSONAL DATA AND RECOMMENDER SYSTEM

From a legal perspective, Recital 12 of the Artificial Intelligence Act is particularly relevant because it identifies the distinctive feature of AI systems in their capacity to infer outputs from inputs or data. These outputs may include predictions, content, recommendations, or decisions capable of influencing physical or virtual environments. This is central to the present analysis, as it shows that AI systems do not merely store or process data, but use data to produce outputs that can shape users' online experiences and support decision-making processes. Moreover, Recital 12 clarifies that this inferential capacity may derive from machine learning techniques, which are learned from data, or from logic- and knowledge-based approaches. In this sense, the legal definition of AI directly connects data processing with prediction, recommendation, and decision-making.[30] In this context, personal data may play different roles within AI systems. First, they may serve as an input to the operation of the system, for example where a recommender

29 Article 29 Data Protection Working Party, Guidelines on Profiling, WP251rev.01, 2018
30 Regulation (EU) 2024/1689 (Artificial Intelligence Act), Recital 12 and Article 3(1).

system uses information about a user's interests to identify content that may be relevant to that user. Secondly, personal data may also be produced as an output, where the AI system generates new information concerning an individual, such as a risk score or assessment. Finally, personal data may function as a building block for the AI model itself, since machine learning systems are often trained on datasets containing information about individuals. This distinction is useful because it shows that the relationship between AI and personal data is not limited to collection: personal data may be used to train the model, operate the system, and generate new information about individuals.[31] A specific form of AI particularly relevant in this context is predictive artificial intelligence. Its functioning depends on the availability of large quantities of data, since reliable predictions cannot be produced without adequate datasets. In practice, predictive AI is used in several contexts, such as fraud detection, recommendation systems, churn prediction, preventive maintenance, and logistics optimization. However, predictive AI does not produce certain results, but predictions and hypotheses that must be assessed in relation to the specific context in which they are used. For this reason, the accuracy and reliability of predictive AI depend on the quality and quantity of training data, as well as on data cleaning, validation, and regular dataset updates. At the same time, attention must be paid to ethical concerns and potential biases, to avoid unfair or unreliable outcomes.[32] AI-driven personalization represents one

31 ALMADA M., Law & Compliance in AI Security & Data Protection, Training curriculum on AI and data protection, Support Pool of Experts Programme, European Data Protection Board, 2024.
32 Alalaq, A. S. (2026). Perdictive Artificial Intelligence: A theoretical analytical study of foundations and concepts. H2D | Revista de Humanidades Digitais, 8.

of the main applications of predictive AI within digital platforms. By using machine learning and deep learning techniques, platforms can infer users' interests, predict behaviors and dynamically adapt content, interfaces, and interactions to individual preferences. Recommendation systems, including collaborative filtering and content-based approaches, are used to match user preferences with relevant products, content or services, increasing engagement, satisfaction, and conversion rates. However, the same mechanisms raise concerns when personalization becomes intrusive or opaque, as users may lose trust and be exposed to filter bubbles or echo chambers that limit the diversity of information they receive.[33]In this context, recommender systems used by dominant social media platforms can be understood as a prominent example of what has been described as 'high reach AI'. High-reach AI refers to broadly used data-driven AI systems whose widespread adoption may generate significant risks for individual users over time and, in aggregate, for society. These systems are particularly relevant because they are not marginal technologies but have become an integral part of everyday internet experience and are used by the largest social media platforms. In the European regulatory context, the Digital Services Act defines a recommender system as a fully or partially automated system used by an online platform to suggest specific information to users or to determine the relative order or prominence of information displayed. Recommender systems therefore go

33 Nwanna M, Offiong E, Ogidan T, et al. AI-Driven Personalisation: Transforming User Experience Across Mobile Applications. J Artif Intell Mach Learn & Data Sci 2025, 3(1), 1920-1929. DOI: doi.org/10.51219/JAIMLD/maxwell-nwanna/425

beyond simple ranking, as they use user data to adapt and personalize suggestions over time. This may affect the information users' access and, consequently, their online behavior. The relevance of these systems is also linked to the fact that they are often designed to optimize platform-level performance indicators, such as engagement, traffic, advertising revenue, or the accuracy of recommendations. While this may provide users with more relevant content and reduce information overload, it may also produce unintended personal and social effects, especially where users are steered towards content that is not necessarily beneficial to them or to society in the long term. A key legal issue is the opacity of these systems. The operational opacity maintained by social media platforms makes it difficult to assess the societal impact of recommender systems, particularly because their large-scale effects are not easily observable from the perspective of individual users. For this reason, the DSA's transparency provisions become central.[34] Article 27 of the DSA requires providers of online platforms using recommender systems to indicate, in plain and intelligible language, the main parameters used by those systems and any options available to users to modify or influence them. These parameters must explain why certain information is suggested. In this way, Article 27 addresses the opacity of recommender systems by requiring platforms to provide users with clearer information about how content is selected, ranked, or prioritized.[35] For very large online platforms and very large online search engines,

34 Söderlund, K., Engström, E., Haresamudram, K., Larsson, S., & Strimling, P. (2024). Regulating high-reach AI: On transparency directions in the Digital Services Act. Internet Policy Review, 13(1). https://doi.org/10.14763/2024.1.1746

35 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), Article 27.

Article 38 introduces an additional safeguard, providers of very large online platforms and very large online search engines using recommender systems must to ‘provide at least one option for each of their recommender systems which is not based on profiling’.36 The relevance of this requirement lies in the impact that recommender systems may have on users. As highlighted by the DSA, such systems can significantly affect the ability of recipients to retrieve and interact with information online and may play an important role in the amplification of certain messages, the viral dissemination of information and the stimulation of online behaviour. For this reason, the non-profiling option aims to ensure that users are not necessarily subject to recommendations based on the profiling of their personal data. In this regard, the EDPB stresses that the available options should be presented equally and that users should not be nudged towards the profiling-based system. Moreover, when the non-profiling option is active, providers should not continue collecting and processing personal data to profile users for future recommendations.[37]

36 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), Article 38.
37 European Data Protection Board, Guidelines 3/2025 on the interplay between the DSA and the GDPR, Version 1.1, adopted on 11 September 2025.

# LEGAL LIMITS TO DATA-DRIVEN AI AND PLATFORM PRACTICES

## 3.1 GDPR PRINCIPLE AS LIMITS TO DATA-DRIVEN AI

Article 5(1)(b) GDPR establishes the principle of purpose limitation, according to which personal data must be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. This principle is particularly relevant in the context of AI and big data technologies, which often enable the reuse of personal data for purposes different from those for which the data were originally collected. For example, data initially collected for contract management may later be used to infer consumer preferences or support targeted advertising. The legitimacy of such repurposing depends on whether the new purpose is compatible with the original one. According to the Article 29 Working Party, relevant factors include the relationship between the original and the new purpose, the expectations of the data subject, the nature of the data, the possible impact on individuals, and the safeguards adopted by the controller. In the AI context, a distinction can be drawn between the use of personal data for training a model and the use of personal data as input for generating inferences about a specific individual. While the inclusion of personal data in a training dataset may not directly affect the individual concerned, the use of personal data to generate individual predictions or assessments has a direct impact on the data subject and therefore requires a stricter application of the purpose limitation principle. Closely linked to purpose limitation is the principle of data minimisation under Article 5(1)(c) GDPR, according to which personal data must

be adequate, relevant and limited to what is necessary for the purposes of the processing. This principle creates a tension with AI and big data analytics, since these technologies often rely on large datasets to identify new and unexpected correlations. This tension can be reduced by reading data minimisation together with proportionality. The use of additional personal data is not automatically excluded, if it brings a benefit in relation to the purpose pursued and does not create disproportionate risks for data subjects. In this regard, safeguards such as pseudonymisation and appropriate security measures may help reduce the risks connected to the retention and use of personal data. A distinction should also be made between statistical processing and individualised inferences. Where personal data are used only to produce aggregate statistical results, and not to take decisions concerning specific individuals, data minimisation may be applied less strictly. By contrast, where data are used to assign a person to a group or to infer information about that individual, the resulting information remains personal data and falls within the scope of data protection. [38] Another relevant limit concerns transparency, particularly where AI systems are used to support or make automated decisions. The transparency obligations established by the GDPR have nevertheless generated a significant academic debate regarding the existence of a so-called 'right to explanation' in relation to automated decision-making. According to Wachter, Mittelstadt and Floridi, the GDPR does not actually provide

38 European Parliamentary Research Service (EPRS), *The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence*, Panel for the Future of Science and Technology (STOA), PE 641.530, June 2020.

individuals with a general legal right to obtain an explanation of a specific automated decision. Rather, the Regulation requires controllers to provide meaningful information about the logic involved in automated processing, as well as information concerning its significance and envisaged consequences for the data subject. The authors argue that Articles 13, 14 and 15 GDPR establish what may be described as a 'right to be informed', rather than a comprehensive right to explanation. Furthermore, Article 22 GDPR, which regulates decisions based solely on automated processing, grants individuals the right to obtain human intervention, express their point of view and contest a decision, but it does not explicitly require controllers to explain how a specific decision was reached. The absence of a clear and legally binding right to explanation has been interpreted as a significant limitation of the GDPR framework, particularly in the context of complex AI systems whose decision-making processes are often difficult to understand and scrutinise. As a result, although the GDPR promotes transparency as a fundamental principle, important challenges remain regarding the practical explainability and accountability of algorithmic decision-making.[39]

### 3.2 PROFILING, AUTOMATED DECISION-MAKING AND MANIPULATIVE PRACTICE

While the previous chapter examined profiling mainly as a data-driven practice within digital platforms, profiling also becomes legally relevant when its results

39 Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law* 76.

are used to guide decisions or actions concerning individuals. In this sense, profiling does not only involve the analysis of personal data, but also the categorisation of individuals according to inferred characteristics. In the context of law enforcement and border management, profiling may be used for two main purposes: first, to identify individuals on the basis of specific intelligence concerning a particular person or event; secondly, as a predictive method to identify unknown individuals who may be of interest to the competent authorities. This second use is particularly significant because it relies on data analysis and informed assumptions. Although predictive methods should ideally focus on behaviour, in practice they may also rely on visible physical characteristics, such as age, gender or ethnicity. This raises concerns because profiling may establish general correlations that do not necessarily hold true for every individual. Profiles may also generate incorrect correlations concerning individuals or groups. Even where certain assumptions reflect a statistical basis, they remain problematic when they lead to individuals being treated primarily as members of a group rather than according to their individual situation.[40] Profiling becomes particularly relevant when the information it generates is used to make or significantly inform decisions about individuals. In this sense, profiling does not merely produce descriptive knowledge but may become part of decision-making processes carried out with different degrees of human intervention and automation. Beyond formal decisions, profiling may also be used to personalise an individual's environment. In online

40 European Union Agency for Fundamental Rights (FRA), *Preventing Unlawful Profiling Today and in the Future: A Guide*, Publications Office of the European Union, Luxembourg, 2018

contexts, profiling is used to automatically personalise experiences and information exposure in real time, directing information towards an individual's presumed interests. Such automated decisions may also be based on a person's predicted vulnerability to persuasion or inferred purchasing power. Examples include the ordering of content in social media news feeds, personalised search results in e-commerce platforms, and the modification of choice architectures in smart environments. Profiling becomes particularly problematic when it is used to inform or feed into decisions affecting individuals, since inaccurate, inferred or biased profiles may lead to misidentification, misclassification or misjudgement. [41] Article 22 GDPR regulates decisions based solely on automated processing, including profiling, where such decisions produce legal effects concerning the data subject or similarly significantly affect them. According to the Article 29 Working Party, this provision should be understood as establishing a general prohibition on fully automated individual decision-making, rather than a right that applies only when actively invoked by the data subject. Consequently, individuals are protected from the potential effects of such processing regardless of whether they take any action. A central requirement of Article 22 is that the decision must be based solely on automated processing. This means that there is no human involvement in the decision-making process. The mere formal participation of a human actor is not sufficient to exclude the application of Article 22. Human involvement must be meaningful and exercised by a person with the authority and competence to assess

41 Frederike Kaltheuner and Elettra Bietti, 'Data is Power: Towards Additional Guidance on Profiling and Automated Decision-Making in the GDPR' (2018) 2(2) *Journal of Information Rights, Policy and Practice*

the relevant information and, where appropriate, modify the outcome of the decision. The GDPR nevertheless allows automated decision-making in specific circumstances, namely where the decision is necessary for entering into or performing a contract, is authorised by Union or Member State law, or is based on the explicit consent of the data subject. In such cases, appropriate safeguards must be implemented to protect the rights, freedoms and legitimate interests of individuals.[42] After considering the limits imposed by Article 22 GDPR on solely automated decision-making, it is possible to move to the AI Act, which addresses certain AI practices considered unacceptable because of their potential impact on individuals' autonomy and behaviour. The analysis may turn to Article 5(1)(a) and (b) of the AI Act, which concerns harmful manipulation, deception and exploitation. These prohibitions aim to protect individuals and vulnerable persons from AI-enabled practices that may materially influence their behaviour. Article 5(1)(a) prohibits AI systems that deploy subliminal techniques beyond a person's consciousness, or purposefully manipulative or deceptive techniques, where these techniques materially distort behaviour by impairing the person's ability to make an informed decision and cause, or are reasonably likely to cause, significant harm. Article 5(1)(b) concerns AI systems that exploit vulnerabilities due to age, disability, or a specific social or economic situation. The rationale of these

42 Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, WP251rev.01, adopted on 3 October 2017, as last revised and adopted on 6 February 2018.

prohibitions is to protect individual autonomy, decision-making and free choice from manipulative, deceptive or exploitative AI practices.[43]

### 3.3 THE EUROPEAN REGULATORY RESPONSE: GDPR, DSA, DMA AND AI ACT

The increasing importance of digital platforms, the expansion of the data economy and the growing integration of artificial intelligence into digital services have raised challenges that traditional legal frameworks have struggled to address. According to recent literature, these developments have required the European Union to rethink its regulatory approach to respond not only to issues of market power and competition, but also to concerns relating to fundamental rights, data protection, transparency and user protection. The European response has progressively taken the form of a coordinated regulatory framework composed of the GDPR, the Digital Services Act (DSA), the Digital Markets Act (DMA) and the Artificial Intelligence Act (AI Act). Rather than constituting isolated legislative interventions, these instruments can be understood as complementary elements of a broader regulatory strategy designed to govern the ecosystem-based nature of contemporary digital markets. This approach reflects the recognition that digital markets no longer operate through separate and independent sectors. Instead, digital platforms, data processing activities and AI systems are increasingly interconnected and organised according to a network logic, creating complex digital ecosystems characterised by network effects, large-scale data collection and

43 European Commission, Communication from the Commission: Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act), C(2025) 884 final, 4 February 2025.

algorithmic decision-making. In this context, the challenges posed by digitalisation cannot be effectively addressed through a single legal instrument but require a coordinated regulatory response capable of tackling different dimensions of the same phenomenon. Within this coordinated framework, each regulation performs a specific function. The GDPR harmonises data protection rules and guarantees the rights to privacy and personal data protection, including through limits on automated decision-making. The DSA establishes the responsibilities of online platforms and strengthens transparency in relation to content moderation, algorithmic ranking and users' rights to justification and redress. The DMA regulates the market behaviour of gatekeepers, contributing to fair competition, consumer choice and economic pluralism. Finally, the AI Act addresses AI-related risks by prohibiting manipulative or discriminatory practices and by imposing strict requirements on high-risk AI systems, including transparency and the quality of training datasets. Together, these instruments define the basic legal framework for the functioning of the digital space and reflect a common human-centred approach based on the protection of fundamental rights, transparency and risk prevention. [44] A particularly significant aspect of the European regulatory framework concerns the relationship between the GDPR and the AI Act. The latter expressly recognises the continued applicability of Union data protection law, providing that it is without prejudice to existing rules on the protection of personal data. As a result, whenever an AI system processes personal data, the GDPR

44 Kitti Mezei, 'Governing Digital Ecosystems in the EU: A Coordinated Regulatory Approach' (2025) 1 *DOT.PL* 1

remains the primary legal framework governing such processing. Despite this complementarity, the two instruments are based on different regulatory approaches. The GDPR adopts a horizontal and principle-based model that applies to all personal data processing activities, regardless of the sector or technology involved. Its framework is built around principles such as lawfulness, fairness, transparency, purpose limitation, data minimisation and accountability. The AI Act, by contrast, focuses specifically on AI systems and introduces obligations linked to their design, development and deployment. According to the European Parliament study, the interaction between the GDPR and the AI Act may generate different forms of regulatory interplay. In some cases, the two frameworks overlap by imposing comparable requirements, such as transparency obligations, ex ante assessments and security-related measures. In other situations, inconsistencies may emerge where obligations diverge or require different regulatory approaches. Finally, certain areas may remain insufficiently clarified, creating gaps regarding the practical reconciliation of the two legal regimes.[45] The broader European regulatory framework also includes the Digital Services Act (DSA) and the Digital Markets Act (DMA), which address dimensions of the digital ecosystem that go beyond the protection of personal data. The DSA focuses on the governance of online platforms and intermediary services, introducing obligations aimed at creating a safer, more predictable and trustworthy online environment. It

45 Hans Graux, Krzysztof Garstka, Nayana Murali, Jonathan Cave and Maarten Botterman, *Interplay between the AI Act and the EU Digital Legislative Framework*, European Parliament, Policy Department for Transformation, Innovation and Health, PE 778.575, October 2025.

establishes differentiated responsibilities depending on the type and size of the service provider and imposes enhanced obligations on very large online platforms and very large online search engines. The regulation is designed to strengthen transparency, accountability and the protection of fundamental rights in the digital environment. The DMA pursues a different objective. Rather than regulating content or platform governance, it focuses on the market power of large digital platforms designated as gatekeepers. The regulation seeks to ensure fairness and contestability in digital markets by imposing a set of obligations and prohibitions on gatekeepers whose position allows them to control access to important digital services and ecosystems. In this way, the DMA addresses concerns relating to economic concentration, data-driven market power and barriers to competition. Together with the GDPR and the AI Act, the DSA and the DMA contribute to a multi-layered regulatory framework in which different instruments address distinct but interconnected aspects of the digital economy. While the GDPR regulates the processing of personal data, the DSA governs the responsibilities of online platforms, the DMA addresses market power in digital ecosystems, and the AI Act focuses on the risks associated with artificial intelligence systems.[46]

46 Cristina Schepisi, Valeria Capuano and Sarah Lattanzi (eds), *An Introduction to Digital Data Law in the EU: Regulatory Framework and Future Perspectives*, Giappichelli, Turin, 2025

# CASE STUDY CLEARVIEW AI AND THE LEGAL LIMITS OF BIOMETRIC DATA EXPLOITATION

## 4.1 FACTUAL BACKGROUND OF THE CASE AND BUSINESS MODEL FOR CLEARVIEW

In January 2020, journalist Kashmir Hill brought public attention to Clearview AI and its facial recognition technology.[47]Clearview AI is a facial recognition technology company whose activities attracted significant public attention following reports concerning its large-scale collection of photographs from the internet. According to available information, the company developed its facial recognition technology using more than three billion photographs scraped from publicly accessible websites, including social media platforms such as Facebook, Twitter and YouTube. The resulting database was reported to be substantially larger than that used by the Federal Bureau of Investigation (FBI).The company offers a facial recognition search engine that enables the identification of individuals by comparing an uploaded image with the photographs contained in its database. The technology has been used by government agencies and military authorities for purposes including intelligence gathering, law enforcement and national security.[48]The functioning of the system is based on the collection and aggregation of photographs obtained from online sources. When a user uploads a portrait of an individual to the application, the system searches the database and returns matching images together with links to the websites from which the

47 Camilla Dul, 'Facial Recognition Technology vs Privacy: The Case of Clearview AI' (2022) 2022 Queen Mary Law Journal 1

48 International Conference on Theory and Practice of Electronic Governance (ICEGOV 2024), October 01–04, 2024, Pretoria, South Africa. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3680127.3680200

photographs were originally scraped. Through this process, the application enables the identification of individuals by connecting facial images with information available online. Initially, Clearview stated that its facial recognition application was sold exclusively to law enforcement authorities in the United States and Canada. However, subsequent reports indicated that the technology had also been used by private companies and private individuals. The company's business model therefore relied on the large-scale collection of publicly available photographs and their incorporation into a searchable facial recognition database. According to the literature, the combination of facial images with other online information allows the identity of individuals to become accessible to users of the application, often without the awareness of the persons concerned. For this reason, the Clearview AI case rapidly became a prominent example in discussions concerning facial recognition technologies, privacy and personal data protection in the digital environment.[49] According to the Italian Data Protection Authority, Clearview's platform cannot be equated with an ordinary search engine. The company does not merely collect images and make them accessible to clients; rather, it processes images obtained through web scraping by means of biometric techniques and a proprietary facial matching algorithm. The collected images are transformed into vector representations and biometric templates, which are then indexed and compared with the image submitted by the user through a one-to-many search

49 Camilla Dul, 'Facial Recognition Technology vs Privacy: The Case of Clearview AI' (2022) 2022 QMLJ 1

process. The service is therefore designed to generate high-quality investigation leads. The results returned by the system may also include metadata associated with the images, such as the title of the image or webpage, the source link, geolocation, gender, date of birth, nationality and language. Moreover, an image collected by Clearview may remain in the database even if the original photograph or the webpage from which it was taken is later removed or made private.[50]

### 4.2 BIOMETRIC DATA AND THE QUALIFICATION AS SPECIAL CATEGORY DATA UNDER EU LAW

The GDPR provides a definition of biometric data in Article 4(14). According to the Regulation, biometric data are 'personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopy data'. [51] Facial recognition technology (FRT) falls within the broader category of biometric technologies. According to the European Data Protection Board (EDPB), biometrics include automated processes used to recognise individuals by quantifying physical, physiological or behavioural characteristics. These characteristics constitute biometric data because they allow or confirm the unique identification of a person. In the context of facial recognition, the processing begins with the collection of a

[50] Garante per la Protezione dei Dati Personali, *Order against Clearview AI* (Order No 50, 10 February 2022) Doc Web No 9751362.
51 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [2016] OJ L119/1, art 4(14).

facial image, referred to as a biometric sample. Through specific technical processing, the system extracts a digital representation of the distinctive features of the face, known as a biometric template. This template may be stored in a biometric database and is intended to be unique and specific to an individual. During the recognition phase, the template is compared with other templates previously generated from facial images. Facial recognition therefore involves a two-step process: the collection of a facial image and its transformation into a biometric template, followed by the comparison of that template with one or more other templates. The EDPB further explains that facial recognition may fulfil two distinct functions. The first is authentication, which aims to verify that a person is who he or she claims to be through a comparison between two biometric templates. The second is identification, which aims to find a person within a group of individuals, an image or a database by comparing one template against multiple templates. In both cases, facial recognition relies on the processing of biometric data relating to an identified or identifiable natural person. Importantly, the EDPB clarifies that a simple photograph does not automatically constitute a facial recognition system. Photographs must undergo specific technical processing in order to extract biometric data. For this reason, facial recognition technologies differ from ordinary image collection or storage systems, since they involve the creation and comparison of biometric templates for the purpose of recognising individuals. Both authentication and identification involve the processing of biometric data and therefore constitute the processing of personal data. More

specifically, the EDPB states that such processing concerns special categories of personal data.[52] Article 9 GDPR regulates the processing of special categories of personal data. Under Article 9(1), the processing of certain categories of personal data is prohibited, including biometric data processed for the purpose of uniquely identifying a natural person. This provision places biometric data within a stricter legal regime when they are used for unique identification. The prohibition laid down in Article 9(1) is not absolute. Article 9(2) provides specific exceptions, including the explicit consent of the data subject, the protection of vital interests, processing relating to data manifestly made public by the data subject, the establishment, exercise or defence of legal claims, reasons of substantial public interest, public health, or archiving, scientific, historical research and statistical purposes, where the conditions required by the Regulation are met. Article 9 also allows Member States to maintain or introduce further conditions, including limitations, regarding the processing of genetic data, biometric data or data concerning health.[53]

52 European Data Protection Board, *Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement* (Version 2.0, adopted 26 April 2023).
53 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [2016] OJ L119/1, art 9.

### 4.3 LEGAL ASSESSMENT AND REGULATORY RESPONSE

At the outset, Clearview AI contested the applicability of the GDPR and the jurisdiction of the Italian Data Protection Authority. The company argued that it was not established in the European Union, did not offer services in Italy or Europe, and had adopted technical measures to block access from European IP addresses. It also denied that its activity amounted to monitoring under Article 3(2)(b) GDPR, claiming that its software merely returned image search results and was comparable to Google Search.[54] Article 3(2) GDPR provides that the Regulation also applies to controllers not established in the European Union where they offer goods or services to individuals in the Union or monitor their behaviour within the Union.[55]The Garante rejected this position. Since Clearview had no establishment in the European Union, the Authority assessed the case under Article 3(2) GDPR. It found that Clearview processed data relating to individuals located in the Union and in Italy, since the complainants' images had been collected, associated with metadata and subjected to biometric processing. The Authority also noted that Clearview had previously made its service available to European users and had later closed European accounts and blocked European IP addresses. The Garante further held that Clearview's activity fell within Article 3(2)(b) GDPR because it was connected to the monitoring of individuals in the Union. The service did not

[54] Garante per la Protezione dei Dati Personali, Ordinanza ingiunzione nei confronti di Clearview AI (Provvedimento n 50, 10 February 2022) doc web n 9751362.
[55] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [2016] OJ L119/1, art 3.

operate as a simple search engine: it collected online images, transformed them into biometric templates, compared them with images uploaded by users and extracted associated information. For this reason, the Authority concluded that the GDPR was applicable to Clearview's processing activities. After establishing the applicability of the GDPR, the Garante assessed the lawfulness of Clearview's processing activities. First, it found a violation of Article 5(1)(a), (b) and (e) GDPR.[56] Article 5 GDPR establishes the main principles governing the processing of personal data. Under points (a), (b) and (e), data must be processed lawfully, fairly and transparently, collected only for specific, explicit and legitimate purposes, and stored in an identifiable form only for as long as necessary to achieve those purposes.[57]According to the Authority, the processing did not comply with the principles of lawfulness, fairness and transparency, since the data subjects had no direct contact with Clearview and were not adequately informed about the processing. The purpose limitation principle was also infringed, because the public availability of online images did not allow Clearview to reuse them freely for facial recognition purposes. In addition, the Garante found a breach of the storage limitation principle, as Clearview had not indicated any retention period and the images appeared to remain in its database even when they had been removed or made private at the original source. The Garante then examined Article 6 GDPR.

[56] Garante, Clearview AI (n 53)
[57] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, Art. 5, 'Principles relating to processing of personal data'.

[58]Article 6 GDPR establishes the legal bases that make the processing of personal data lawful.[59]The Garante concluded that Clearview had no valid legal basis for the processing. The company had not obtained the consent of the data subjects, and the legitimate interest invoked by Clearview could not prevail over the rights and freedoms of the individuals concerned, given the intrusive nature of the processing, the biometric elaboration of the images and the large number of persons involved. For the same reasons, the Authority also found a violation of Article 9 GDPR. [60] Article 9 GDPR prohibits the processing of special categories of personal data, including biometric data used for unique identification, unless one of the specific exceptions listed in Article 9(2) applies.[61] Clearview did not merely collect ordinary images but transformed them into biometric data through technical processing. Since biometric data are subject to a stricter legal regime, the Garante held that Clearview could not rely only on Article 6 GDPR and that no valid exception under Article 9 applied. Finally, the Garante found violations of Articles 12, 13, 14 and 15 GDPR. The responses given to access requests were incomplete or delayed, and in some cases, Clearview required excessive identification documents. The privacy policy was also considered insufficient, since it did not provide complete information on essential aspects of the processing, such as the legal basis and the retention period. Since Article 3(2) GDPR was applicable, the Authority further held that Clearview

[58] Garante, Clearview AI (n 53)
[59] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, Art. 6, 'Lawfulness of processing'.
[60] Garante, Clearview AI (n 53)
[61] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [2016] OJ L119/1, art 9.

was required to designate a representative in the European Union under Article 27 GDPR.[62]Article 27 GDPR requires non-EU controllers falling within the territorial scope of the GDPR to appoint a representative in the European Union for compliance and communication purposes.[63]The failure to do so constituted an additional violation. The Garante concluded that Clearview AI had unlawfully processed personal and biometric data in violation of several provisions of the GDPR. As a result, it prohibited the continuation of the processing activities concerning individuals located in Italy, ordered the deletion of the related data, and imposed an administrative fine of €20 million.[64]

[62] Garante, Clearview AI (n 53)
[63] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [2016] OJ L119/1, art 27
[64] Garante, Clearview AI (n 53)

# CONCLUSIONS

This thesis has shown that personal data play a central role in the contemporary digital economy. They are collected and processed by digital platforms and AI systems to improve services, personalize content, support recommender systems and generate advertising revenues. Their value does not depend simply on their existence, but on the ability of firms to organize, analyse and transform them into useful and monetizable information. At the same time, the economic exploitation of personal data raises significant legal concerns. Profiling, automated decision-making and AI-driven personalization may affect individuals' choices, opportunities and access to information. These practices can create risks of opacity, discrimination, manipulation and loss of control over personal data. The European legal framework responds to these risks through a combination of instruments. The GDPR establishes general principles for data processing, such as lawfulness, fairness, transparency, purpose limitation and data minimisation. The Digital Services Act introduces specific transparency obligations for online platforms and recommender systems. The Digital Markets Act addresses the power of large digital gatekeepers, while the Artificial Intelligence Act introduces a risk-based approach to AI systems and prohibits certain manipulative or exploitative practices. The Clearview AI case confirms the importance of these safeguards. The company's large-scale collection of facial images from the internet and their transformation into biometric templates show that publicly available data cannot automatically be used without limits. Biometric data require protection because

they allow the unique identification of individuals and are closely linked to personal identity. In conclusion, personal data are both an economic resource and a fundamental rights issue. Their use can support innovation and digital services, but it must remain compatible with transparency, accountability and the protection of individuals. The main challenge for the future is to ensure that the digital economy does not develop only through the extraction of data, but through a model that respects individual rights and places the person at the centre of technological progress.

## REFERENCES


ALALAQ A. S., Predictive Artificial Intelligence: A Theoretical Analytical Study of Foundations and Concepts, 'H2D | Revista de Humanidades Digitais', vol. 8, 2026.

ALMADA M., Law & Compliance in AI Security & Data Protection, European Data Protection Board, Brussels, 2024.

ARTICLE 29 DATA PROTECTION WORKING PARTY, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation (EU) 2016/679, WP251rev.01, 2018.

BATAINEH A. S., MIZOUNI R., EL BARACHI M., BENTAHAR J., Monetizing Personal Data: A Two-Sided Market Approach, 'Procedia Computer Science', vol. 83, 2016, pp. 472-479.

BIRCH K., COCHRANE D., WARD C., Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data, 'Big Data & Society', 2021.

CIFALDI G., Evolution of Concepts of Privacy and Personal Data Protection under the Influence of Information Technology Development, 'Sociology and Social Work Review', vol. 7, n. 1, 2023, pp. 35-60.

CRÉMER J., DE MONTOYA Y.-A., SCHWEITZER H., Competition Policy for the Digital Era, Publications Office of the European Union, Luxembourg, 2019.

DUL C., Facial Recognition Technology vs Privacy: The Case of Clearview AI, 'Queen Mary Law Journal', 2022.

EUROPEAN COMMISSION, Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act), C(2025) 884 final, Brussels, 2025.

EUROPEAN DATA PROTECTION BOARD, Guidelines 05/2020 on Consent under Regulation 2016/679, Brussels, 2020.

EUROPEAN DATA PROTECTION BOARD, Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement, Version 2.0, Brussels, 2023.

EUROPEAN DATA PROTECTION BOARD, Guidelines 3/2025 on the Interplay between the DSA and the GDPR, Version 1.1, Brussels, 2025.

EUROPEAN DATA PROTECTION SUPERVISOR, Opinion 3/2018 on Online Manipulation and Personal Data, Brussels, 2018.

EUROPEAN PARLIAMENT AND COUNCIL, Regulation (EU) 2016/679 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 (General Data Protection Regulation), Official Journal of the European Union, L 119, 2016.

EUROPEAN PARLIAMENT AND COUNCIL, Regulation (EU) 2022/1925 on Contestable and Fair Markets in the Digital Sector (Digital Markets Act), Official Journal of the European Union, L 265, 2022.

EUROPEAN PARLIAMENT AND COUNCIL, Regulation (EU) 2022/2065 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), Official Journal of the European Union, L 277, 2022.

EUROPEAN PARLIAMENTARY RESEARCH SERVICE, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, STOA, PE 641.530, Brussels, 2020.

EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, Preventing Unlawful Profiling Today and in the Future: A Guide, Publications Office of the European Union, Luxembourg, 2018.

FARINHO D. S., Personal Data Processing by Online Platforms and Search Engines: The Case of the EU Digital Services Act, 'Public Governance, Administration and Finances Law Review', vol. 9, n. 1, 2024, pp. 37-58.

FEDERICI M., GIULIANO S., Italy's Tax Authorities Target 'Free' Online Services for VAT, 'Bloomberg Tax', 13 June 2025.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Order against Clearview AI (Order No. 50 of 10 February 2022), Doc. Web No. 9751362, 2022.

GRAUX H., GARSTKA K., MURALI N., CAVE J., BOTTERMAN M., Interplay between the AI Act and the EU Digital Legislative Framework, European Parliament, Brussels, 2025.

KALTHEUNER F., BIETTI E., Data is Power: Towards Additional Guidance on Profiling and Automated Decision-Making in the GDPR, 'Journal of Information Rights, Policy and Practice', vol. 2, n. 2, 2018.

LERNER A. V., The Role of 'Big Data' in Online Platform Competition, 'SSRN Electronic Journal', 2014, pp. 7-20.

MEZEI K., Governing Digital Ecosystems in the EU: A Coordinated Regulatory Approach, 'DOT.PL', n. 1, 2025.

MURSIA M., TROVATO C. A., The Commodification of Our Digital Identity: Limits on Monetising Personal Data in the European Context, 'Rivista di Diritto dei Media', 2021.

NWANNA M., OFFIONG E., OGIDAN T. et al., AI-Driven Personalisation: Transforming User Experience Across Mobile Applications, 'Journal of Artificial Intelligence, Machine Learning and Data Science', vol. 3, n. 1, 2025, pp. 1920-1929.

OECD, Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value, OECD Publishing, Paris, 2013.

OECD, Addressing the Tax Challenges of the Digital Economy: Action 1 – 2015 Final Report, OECD Publishing, Paris, 2015.

SCHEPISI C., CAPUANO V., LATTANZI S. (eds), An Introduction to Digital Data Law in the EU: Regulatory Framework and Future Perspectives, Giappichelli, Torino, 2025.

SÖDERLUND K., ENGSTRÖM E., HARESAMUDRAM K., LARSSON S., STRIMLING P., Regulating High-Reach AI: On Transparency Directions in the Digital Services Act, 'Internet Policy Review', vol. 13, n. 1, 2024.

WACHTER S., MITTELSTADT B., FLORIDI L., Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, 'International Data Privacy Law', vol. 7, n. 2, 2017.

ZUBOFF S., The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, PublicAffairs, New York, 2019.