UNIVERSITÀ DEGLI STUDI DEL SANNIO

DIPARTIMENTO DI DIRITTO, ECONOMIA, MANAGEMENT E METODI QUANTITATIVI

Corso Dottorato di Ricerca in "Persona, Mercato, Istituzioni"

XXXVI Ciclo

Tesi di Dottorato in

DIRITTO PRIVATO

(IUS-01 GIUR-01/A)

PROFILING, PERSONALISATION, VULNERABILITY: APPLICATION PARADIGMS. RESEARCH HYPOTHESIS THROUGH THE STUDY OF LARGE LANGUAGE MODELS

TUTOR:

Ch.ma Prof.ssa

ANTONELLA TARTAGLIA POLCINI

CANDIDATA:

Marianna Ciullo

COORDINATORE:

Ch.m Prof.

RICCARDO RESCINITI

Anno Accademico 2023/2024

INDEX

Introductory notes

A phenomenon that has characterised the global set-up since at least the last decade, the study of complexity and complex systems has had a considerable development and impact in any scientific field. Indeed, in recent years mankind has been faced with phenomena of a completely unexpected scope and global dimension: major climate changes, health challenges of even pandemic proportions, major demographic variations at different latitudes and speeds, the crisis of many social structures, the advent and implementation of increasingly sophisticated technological systems, and the almost widespread use of artificial intelligence. Undoubtedly, what has been outlined constitutes a neuralgic turning point not only in the history of mankind but also in the history of human knowledge and human knowledge management. In particular, the so-called *Fourth Industrial Revolution* is considered by some to be the most transformative revolution of any experienced by mankind so far: it succeeded in questioning the very idea of what it means to be 'human'. In other words, the change in the digital sense of interpretation and worldview takes its cue from the informational code of ICTs, which is configured as self-referential. In fact, in order for the digitised system - *rectius* defined from the machine's point of view - to function, the real must be brought back to operational data and machine-processable *inputs*. In this sense, one can understand how the very meaning of being human changes: the identity of the human being finds a *decomposition* followed by a recomposition in the digital ecosphere, possible through the use of data referable to a physical subject. From a legal-philosophical point of view, this complex concept can be traced back to a phrase with a revolutionary effect 'what is real is informational and what is informational is real'. On the basis of these considerations, it is clear how, on a practical level, the influence of Artificial Intelligence[1] has already invested areas such as the labour market, public spaces, as well as and above all private dynamics and dimensions (e.g. to the extent that it is difficult and unthinkable to have an everyday life that is not integrated by the support of this type of technology[2]). Among others, the jurist is called upon to play a role of no small moment in the so-called complexity society. It is interesting, in fact, how many of the great changes mentioned above have repercussions of more than significant scope in the legal sphere. Suffice it to think of the situations of imbalance of power in a world that has now become multipolar; of the

[1]Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS*. IT, (18 of 28/07/2021), pp. V-VI.
[2]Daelman, C., & Yordanova, K. (2023). AI through a human rights lens. The role of human rights in fulfilling AI's potential. Artificial intelligence and the law. Intersentia, Cambridge, 135-168 , p. 123.

uncertainties in the structure and institutional dynamics, sometimes outdated, that find themselves operating in a context that is completely different from the original one; of the unethical use of the means of communication and information, as well as of artificial intelligence, harbingers of computer risks, improper use of data, violations of *privacy*, and highly discriminatory dynamics[3] . In this regard, it is also worth considering how the technological era is characterised by constant evolution: from the network as a mere tool, in a static sense, we have come to the integration of the digital into objects and the replacement of human prerogatives and tasks, realising unprecedented possibilities of interaction between humans and between humans and machines. In this changed context, law is called upon to regulate information technologies and the digital reality by outlining rules, principles and shared values, regulating behaviour, defining responsibilities, and healing conflicts. In performing this function, law is exposed to changes that arise from the characteristics and evolution of the aforementioned context. In fact, the object of regulation, today constituted by the so-called *intangible goods*, changes; the context of legislative intervention changes, since man is called upon to intervene through legal rules on the digital and robotic reality in order to guarantee the protection of rights; the temporal dimension changes, if technological evolution proceeds with extreme rapidity, the law is structurally characterised by different timeframes. Another major change concerns the spatial dimension: the digital space, except for a few cases, stands as a decentralised and aterritorial public space, over which no one can claim exclusive power, so that the applicable law must also be rethought. Several technological giants have long since conquered the global dimension, dictating rules that apply to individuals. Among other things, the technological revolution has had the power to define new geometries of power, blurring the boundary between the public and private dimensions of special interests. The role of law and the protections afforded in the area of non-discrimination appear crucial. On this point, consider that the commercialisation of AI systems has posed and continues to pose various legal and regulatory challenges, especially with regard to fundamental rights and the human person. Since human rights are interdependent and interconnected, AI affects almost all human rights[4] . A first category of human rights concerns the right to life[5] , liberty and

[3] Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J.. & Helbing, D. (2012). Manifesto of computational social science. The European Physical Journal Special Topics, 214, 325-346.

[4] On this point see, more extensively, Daelman, C., & Yordanova, K. (2023). AI through a human rights lens. The role of human rights in fulfilling AI's potential. Artificial intelligence and the law. Intersentia, Cambridge, 135-168.

[5] Article 2 European Convention on Human Rights (ECHR), Article 3 Universal Declaration of Human Rights (UDHR).

security[6] and a fair trial[7] . One of the many applications of AI concerns the use of criminal risk assessment software. Furthermore, a specific application of AI that poses a potential threat to the right to life are fully autonomous weapon systems, more commonly referred to as 'killer robots'. Non-state actors may also have access to this AI technology, which raises concerns as they are not necessarily bound by traditional laws of armed conflict. Given their programmed nature, autonomous weapons may not be able to consider the 'nuances' of a particular situation or react appropriately to unexpected circumstances. This can lead to potentially high error rates with fatal casualties that a human operator would have been able to avoid. Freedom of thought, conscience and religion[8] , as well as freedom of expression[9] and freedom of assembly or association[10] are also fundamental rights. The European Court of Human Rights (ECHR) recognises the important role of the Internet. Undoubtedly, the Internet represents an unprecedented space for the expression of freedom of thought and speech. However, there is an increasing intrusion by artificial intelligence applications that influence the use and dissemination of information online. Search engines and social platforms use such tools to filter and moderate the available content, a process that, paradoxically, may undermine the diversity of voices and opinions. Such algorithmic filtering can also amplify the phenomenon of 'fake news', altering individuals' perceptions and beliefs. The use of Artificial Intelligence for censorship carries a serious risk of restricting religious freedom, as this practice can lead to the selective identification and removal of religious content. Such action not only erodes the individual's right to manifest one's faith, but also undermines the possibility of sharing and discussing such beliefs within the virtual public space. Moreover, AI-powered censorship can threaten freedom of association, as it leads to the deletion of online groups, pages, and content that serve as places of aggregation and exchange for individuals with common interests. This practice not only limits people's ability to organise and collaborate online, but also to actively participate in debate and discussion on issues of collective interest. Another significant manifestation of the ethical implications of artificial intelligence is the surveillance fuelled by this technology. The intersection between the pervasive and imperceptible nature of many artificial intelligence applications and their ability to monitor and identify behaviour increasingly results in self-censorship and censorship, both in the public sphere and even in

[6] Article 5 ECHR, Article 3 UDHR and Article 9 ICCPR.
[7] Article 6 ECHR, Articles 6, 7, 8 and 10 UDHR and Article 14 ICCPR.
[8] Article 9 ECHR, Article 18 UDHR and Article 18 ICCPR.
[9] Article 10 ECHR, Article 19 UDHR and Article 19 ICCPR.
[10] Article 11 ECHR, Article 20 UDHR and Articles 21 and 22 ICCPR.

private communications. Self-censorship emerges when individuals feel constant surveillance of their movements or perceive a lack of anonymity, leading them to modify their behaviour. Therefore, the widespread use by public and private entities of surveillance techniques based on artificial intelligence, such as video surveillance, facial recognition and behavioural analysis, could have a chilling effect on freedom of expression, thought, religion, association and assembly. The application of AI can indeed influence the creation, dissemination and access to information, representing a crucial aspect of ensuring fairness in elections and political participation. AI-based surveillance assumes a particularly relevant role in this context, as it can undermine political participation by identifying, deterring and restricting certain social groups from voting, or by influencing their voting behaviour. This practice risks seriously undermining the integrity and validity of democratic processes by subjecting elections and political participation to potential distortion and manipulation. Failure to address these issues could weaken voter confidence in the legitimacy of elections and democratic political systems in the long run. This scenario, in turn, could threaten the right to effective political participation. Loss of confidence in the integrity and impartiality of electoral processes could erode the very fabric of democracy, discouraging active citizen participation and fuelling political disillusionment and apathy. Therefore, it is crucial to take appropriate measures to address the challenges posed by AI applications in the political sphere in order to preserve and strengthen the foundations of democracy . [11] Undoubtedly, the healthcare sector emerges as one of the fields where applications of Artificial Intelligence promise to bring significant benefits. By analysing and understanding an individual's daily life patterns, AI applications can incentivise healthy behaviour and act as a support in the adoption of more balanced and sustainable lifestyles. In addition, AI proves to be a valuable tool in the early detection of diseases and the formulation of more accurate and timely diagnoses, thus making an essential contribution to improving healthcare and promoting individual and collective wellbeing. The introduction of artificial intelligence and its applications also have significant implications for the protection of intellectual property, highlighting an intense interaction between the two spheres. This interaction can be analysed through three main categories. First, AI itself represents a technology that can be used to facilitate IPR management. This has led to the increasing adoption of AI in the administration of IP protection applications, improving the efficiency and accuracy of management and control processes. Secondly, IP protection provides a

[11]Manheim, K., & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. Yale JL & Tech., 21, 106, pp. 106-188.

legal framework that protects the development and use of AI applications, thus helping to stimulate innovation and creativity in the field of AI. Such legal protection is crucial to ensure that AI developers can fully exploit the value of their creations and invest in further research and development. Thirdly, however, IP rights can also raise questions regarding the need for greater transparency and accountability of AI systems and programmes. As AI becomes increasingly integrated into decisions and processes that affect people's lives, there is a need to balance the protection of intellectual property with ensuring ethical and accountable governance of AI, taking into account the principles of fairness, security and accountability. In general, the progressive recourse to various technical applications to simplify everyday activities entails a massive use of data. The growing and well-known use of systems based on Artificial Intelligence is due not only to the accessibility of increasingly robust yet inexpensive *computational power* and thus of increasingly sophisticated algorithmic models, but also and above all to the accessibility of vast masses of data. Indeed, the more data available for algorithmic training purposes, the greater the degree of algorithmic learning. The race for AI is particularly influenced by the network effects already known from the platform economy: the more users a company has, the more personal data can be collected and processed to train algorithms. This in turn leads to better products and services, which translates into more customers and more data. In light of these network effects, some fear that the market for AI systems will become oligopolistic with high barriers to entry. AI systems challenge the current understanding of privacy, as most AI technologies have a negative impact on the right to privacy. On the one hand, AI systems based on Machine Learning cannot function without data. On the other hand, without AI systems it would not be possible to understand and decipher many of the masses of unstructured data. In other words, personal data are increasingly both the source and the target of AI applications. Consequently, this type of technology creates strong incentives to collect and store as much additional data as possible in order to gain meaningful new insights. This trend is further reinforced by the shift to ubiquitous monitoring and surveillance through *smart* devices and other networked, ubiquitous sensors in the IoT. Artificial intelligence amplifies large-scale surveillance through techniques that analyse video, audio, images and social media content of entire populations. The spread of increasingly sophisticated forms of Artificial Intelligence in everyday life contributes to this development[12] . Furthermore, self-learning algorithms are used by many companies, political parties and other actors to influence and manipulate citizens and consumers

[12] Ebers, M. (2019). Regulating AI and robotics: ethical and legal challenges, p. 23.

through various techniques. This raised the question of how the law could provide adequate safeguards against such practices. Another problem closely linked to algorithmic decision-making is, as stated earlier, the risk of discrimination: many studies indicate, in fact, that algorithms are often not value-neutral, but biased and discriminatory. Here again, one wonders to what extent citizens, consumers and especially the most vulnerable can and should be protected[13] . Another decidedly relevant issue that has affected the legal debate has been, on the basis of what has been mentioned, the sufficiency of the scope of the GDPR with respect to the gradually growing and thus increasingly widespread applications of Artificial Intelligence. On this point, in particular, the lack of data protection rights or obligations relating to Machine Learning models themselves in the period after their construction, but before decisions on their use, has been reflected upon. Also because, as a rule, ML models do not contain strictly personal data, but only information on groups and classes of people[14] . Also for these reasons, the European legislator, first of all on a global scale, sought to implement legislation focusing on Artificial Intelligence from 2021 onwards, legislation that came into being, following a complex, delicate and multifaceted path, on 13 June 2024. In particular, the European Union Regulation laying down rules on Artificial Intelligence, which partially came into force on 1 August 2024, was born with the legislator's intention to provide 'clear requirements and obligations to developers and operators regarding specific uses of AI, while reducing the administrative and financial burdens on businesses'[15] . Starting from such premises, it is worth emphasising how the issues mentioned have important and decisive repercussions in the field of private law: for civil law research, the issues of consent, self-determination, as well as the relative prejudices that might arise and the protections to be provided, are, by way of example, relevant. Moreover, the aforementioned phenomena prove to be of particular importance, also and especially in the field of private law, due to the newly emerging vulnerabilities associated with them. Indeed, the new potentials of machine learning, automated decision-making, personalised behavioural advertising and predictive analytics increase the imbalance of power between individuals and data-processing entities in the digital environment. Of course, the imbalance is sometimes intrinsic to certain legal dynamics, but the European legislator, in regulating the so-called *digital ecosphere*, has focused on the variable level of (dis)equity of this imbalance; in other words, it has focused on the degree to which it affects the autonomy, dignity and

[13] Ebers, M. (2019). Regulating AI and robotics: ethical and legal challenges, p. 29.
[14] Ebers, M. (2019). Regulating AI and robotics: ethical and legal challenges, p. 24.
[15] h ttps://commission.europa.eu/news/ai-act-enters-force-2024-08-01_it

integrity of individuals, who are thus increasingly *vulnerable* .[16] Focusing on a more defined theme than the general framework illustrated, the present research work aims to investigate the profiling activity and its repercussions in the field of jurisprudence, the problem of *clustering*, as well as the practical implications offered by the application of generative artificial intelligence models and *Large Language Models*. In other words, this paper proposes to analyse the legal evolution of the activity of profiling and the automated processing of personal data, with specific reference to the so-called "Large Language Models". *Large Language* Models, taking the vulnerable individual as a reference. The research question that will be attempted to be answered will concern the possible protections to be provided in this sense and the regulatory sufficiency of the existing legislation. For the purposes of this research, the methodology used was characterised by interdisciplinarity and comparison. In addition, the technical-informatics aspects of the research topic were also considered and deepened. In particular, the scope of the research was extended to the problems underlying Large Language Models, a subject that was deepened by obtaining a Level II Master's Degree in "Artificial Intelligence for the Humanities" at the University of Naples "Federico II" during the academic year 2022/2023. With reference to the research topic as a whole, the relevance of the comparative aspect is clear, which was also deepened through a period spent abroad as a Visiting Scholar ("Paris Lodron" Universitaet Salzburg, Department of Law and Economics, Department of Private Law from 9 October 2023 to 9 January 2024).

[16] Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press, p. 1

Chapter I

'Profiling: from EU Regulation 2016/679 to the EU Artificial Intelligence Regulation'.

SUMMARY: 1. Profiling: an analytical and critical approach to terminology and definitions . 1.1. The cultural, legal, economic issues underlying profiling. The advantages of *profiling*. 1.2. Follow-up. The risks underlying the activity of profiling. The legislation protecting personal data. 2.1. Automated processing between opportunity and ambiguity. 2.2. The issue of predictive algorithms.3. The evolution of the regulation of profiling and automated decision-making processes. 3.1. Provisions on profiling before the EU Regulation 2016/679. 3.2. Profiling and automated decision-making in the EU Regulation 2016/679. 3.3. Profiling in subsequent European legislation. 4. Classifications, debated issues, problems underlying profiling. 5. Profiling and technological evolution: the European Union Regulation on Artificial Intelligence.

1. Profiling: an analytical and critical approach to terminology and definitions.

An effective and controversial method used in different sectors of human life and action, profiling plays a major role in several areas: from private interest activities, to those related to the economic and entrepreneurial field, to activities in the public sphere and public interest[17] . Profiling, in fact, offers the possibility of cataloguing individuals in order to predict their behaviour, habits and trends. Through it, it is possible to extract information from data collections and to obtain resources to operate in digital and non-digital environments[18] . In addition, profiling allows access to information that can profile an individual's behaviour over time, as well as using it to predict his or her future behaviour[19] . Indeed, the gathering of information, especially for the benefit of a public power, is a phenomenon that has always existed but has certainly found fertile ground for further implementation and greater accuracy in the new technologies and the so-called *algorithmic society*. Moreover, the use of Artificial Intelligence systems allows for such pervasive and meticulous profiling as to enable predictions so accurate that they come close to certainty[20] . Prior to the advent of EU Regulation 2016/679, various legal literature questioned the definition of profiling. Some legal literature used to define profiling activities as computerised techniques that, through

[17] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.9

[18] Pierucci, A. (2019). Data processing and profiling of individuals. Personal data in European law.

[19] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.10

[20] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.10

data mining in a data archive, enable or aim to enable the classification with a certain probability of a person in a certain category in order to make decisions about that person. According to other parties, however, the activity of profiling would consist of two moments: the first, dedicated to the generation of a profile; the second, concerning the process of dealing with personalities/individuals in the light of the determined profile. In particular, the 2016 legislation of European matrix represents the culmination of a long process, which began with Directive 95/46/EC[21] , which suddenly proved to be ineffective in facing the challenges resulting from the speed of technological evolution and globalisation in the field of personal data protection[22] . Indeed, this directive was subsequently read in the light of the Lisbon Treaty[23] - which reaffirmed the right to the protection of personal data while empowering the European institutions to take legislative action in the field of privacy - but the need to adapt European legislation to the aforementioned rapid developments in technology was strongly felt. The 1995 directive, in any case, is in turn the result of a long jurisprudential history of the Court of Justice of the European Union, which allowed for a pre-trial construction of a right to the protection of personal data that also proved to be a crucial facility to make up for regulatory *shortcomings* and to allow for safeguards that would follow the rapid course of technological evolution[24] . For this reason, the standard is based on the observation that the rapid advance of technological progress, as well as widespread globalisation combined with the concept that sees legal systems to be considered in large macro-areas, have rendered the 1995 directive obsolete in many points. For this reason, the European institutions have designated a single rule, a single rule in the form of a regulation, valid for all Member States and capable not only of fostering economic growth and digital innovation, but also of strengthening the confidence of consumers and users *tout-court* in digital environments and dynamics. Incidentally, a phenomenon that was unforeseeable both

[21] Directive 95/46/EC of the European Parliament and of the Council. Of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (so-called "parent directive").
[22] MONTUORI, L., & SIANO, M. (2017). Evolving the concept of informed consent in the digital world and the transition of traditional marketing to the current challenges of profiling. The new frontiers of privacy in digital technologies: balances and perspectives, 101-126, p. 102.
[23] Article 16 of the Treaty on the Functioning of the European Union (TFEU) , which replaced the former Article 286 of the Treaty establishing the European Community states: '1. Everyone has the right to the protection of personal data concerning him or her. 2. The European Parliament and the Council, acting in accordance with the ordinary legislative procedure, shall lay down the rules relating to the protection of individuals with regard to the processing of personal data by Union institutions, bodies, offices and agencies, and by the Member States when carrying out activities which fall within the scope of Union law, and the rules relating to the free movement of such data. Compliance with these rules shall be subject to control by independent authorities. The rules adopted on the basis of this Article shall be without prejudice to the specific rules laid down in Article 39 of the Treaty on European Union."
[24]ROSSI DAL POZZO, F. (2018). The protection of personal data in the case law of the Court of Justice. EUROJUS, p.5.

for the 1995 legislator and for the *civis*, today's technological evolution has led to a situation whereby a large part of human life and activity takes place on the Net and on the Net: from being an occasional and static place of meeting, work or life, the Net has come to encompass a large part of human activities, also causing entirely new dynamics and new spaces to emerge. Consequently, the traditional orderly arrangements have faded to the point where it is reasonable to reason in large orderly macro-areas (see *above*), which in any case are bound together by an intrinsically homogeneous and borderless digital environment. Naturally, for this reason too, as territorial boundaries are blurred, there are not a few regulatory asymmetries. On this point, therefore, it should be noted that the whole great legislative season on the subject, which opened with the General Data Protection Regulation and continued, through the entry into force of several other acts, up to the European Regulation on Artificial Intelligence, is characterised by the imposition of European rules also on those actors that operate on European soil or with European citizens. In this, one fact is evident, namely that for the most slippery subjects and phenomena that can easily escape human control - above all algorithmic decisions through profiling - the European Union wanted to some extent to offer its citizens greater protection. On the other hand, the GDPR presents similar pillars of law to those considered by Directive 95/46/EC, such as those of information and consent. In fact, EU Regulation 679/2016 revolves around the so-called informed consent, but with novel aspects compared to the past especially with reference to profiling activities[25] . For the aforementioned reasons, the activity under consideration had never been fully defined, also because only in recent decades has it been possible, through different techniques, to arrive at an effective profiling of the user. Therefore, it is Article 4(4) of the GDPR that provides for the first time a complete definition of profiling. Until the advent of the GDPR, the activity of profiling has never been the subject of organic and unified legislation; on the contrary, the regulatory framework relating to this activity has always appeared extremely fragmented and uneven[26] . For instance, Article 15 of Directive 95/46/EC considers automated decision-making processing to be relevant, but only within such processing operations is it possible to find traces of the concept of profiling.

[25] MONTUORI, L., & SIANO, M. (2017). Evolution of the concept of informed consent in the digital world and the transition of traditional marketing to the current challenges of profiling. The new frontiers of privacy in digital technologies: balances and perspectives, 101-126, p.104

[26] Regarding the definition of 'profiling', two visions were opposed in the Italian legal system. Some legal literature used to define profiling activities as computerised techniques that, through data mining in an archive of data, allow or aim to allow a person to be classified with a certain probability in a certain category in order to make decisions on the person. According to another part, however, the activity of profiling would consist of two moments: the first, dedicated to the generation of a profile; the second, concerning the process of dealing with personalities/entities in the light of the determined profile.

Subsequently, in 2002, the so-called e-Privacy Directive (2002/58/EC) intervened, which, as a *lex specialis* with respect to the aforementioned European source, introduced more specific rules on online tracking technologies (such as cookies), as well as the obligation to obtain the user's informed consent for their use for profiling purposes. The Personal Data Protection Code (Legislative Decree 196/2003), on the other hand, connotes the activity of profiling in a negative sense, requiring it to be notified to the Garante per la protezione dei dati personali (Data Protection Authority), since it is included among the risky treatments. Finally, the advent of increasingly complex, innovative technologies capable of creating unprecedented interconnection between users has led to the need to develop new and more appropriate protection strategies. Precisely for these reasons, during the design and drafting phase of the General Data Protection Regulation (GDPR), the need was felt to innovate the regulation of profiling, in order to meet the challenges posed by modern technologies and guarantee more effective protection of individuals' rights. Article 4(4) of the GDPR offers, for the first time, a systematic definition of 'profiling', meaning *'any form of automated processing of personal data consisting of the use of such personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects of that person's professional performance, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements'*. Within the normative assumption, certain references, such as the assessment of aspects of the natural person, as well as the *analysis* and *prediction* of other personal aspects, come to the fore, which prove fundamental to understanding the scope of the rule and its operation in the context of the GDPR. Therefore, it is necessary to read this legal definition in conjunction with Recital 24 and Recital 30. The former, after relying on the territorial scope of application, emphasises the '*monitoring of the data subject's behaviour', whereby* it is appropriate to check whether natural persons are being tracked on the Internet, including the possible subsequent use of data processing techniques (*first and foremost*, the profiling of the individual), in particular to take decisions concerning him or her or to analyse or predict his or her preferences, behaviour and personal attitudes. The second, on the other hand, explicates even further the notion of being tracked online by stating '*Natural persons may be associated with online identifiers produced by the devices, applications, tools and protocols used, such as IP addresses, temporary markers (cookies) or other identifiers such as radio frequency identification tags. Such identifiers can leave traces which, particularly when combined with unique identifiers and other information received by servers, can be used to create profiles of individuals and identify them."* This explains the references to assessment activity, as well as to analysis and prediction in Article 4.

1.1. The cultural, legal, economic issues underlying profiling. The benefits of *profiling*.

Thanks to data sharing systems, to the management and processing of *big data*[27] , as well as to increasingly sophisticated digital techniques, the availability of information is increasing exponentially: this allows profiling to anticipate the user's needs, knowing in advance and in depth his interests. In fact, the phenomenon takes place in the most diverse contexts: for instance, users may be suggested tangible or intangible goods - such as those relating to the entertainment industry -; *social network* users may be profiled in such a way that their *online* social circle expands in a certain direction and forges ties with other, specific users; devices that detect biometric data may suggest to users measures relating to physical or mental health, diet, and other needs relating to psycho-physical well-being. Legal relations, in this case concerning business activity, are profoundly transformed by all these dynamics: in fact, the analysis of data produced by predictive algorithms makes the preferences of the consumer preeminent, who can benefit from having goods and services tailored to his needs. On the other hand, the business activity itself will also benefit from knowing the preferences of its customers in a timely and capillary manner and can increase profit margins[28] . Similarly, also in the relationship with the public administration, an in-depth knowledge of the citizen on the part of the institutions can lead to a more efficient delivery of services, as well as a better allocation of resources[29] . Generally speaking, it can be said that profiling, mainly understood as a mere classification activity, initially takes the form of the creation of groups and profiles. However, when this activity is taken to extremely high and pervasive levels of operation thanks to the adoption of new technologies, a situation is reached in which traditional groups and profiles disappear. In fact, each individual is valued in himself, represented, studied and analysed according to his specific characteristics, tendencies and personal habits. In this context, the issue of the foresight aspect emerges, which constitutes the distinction between classification and profiling according to the Regulation. This predictive element is not limited to the individual in a quasi-micro-economic perspective, but also involves the entire system in a macro-economic perspective. This highlights the real risk of mass surveillance and manipulation, in which predictive analysis and extensive profiling can lead to widespread control of social and economic dynamics.

---

[27]Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big data & society, 1(1), 2053951714528481.

[28] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p. 27.

[29] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.28

1.2. To be continued. The underlying risks of profiling.

While it is true that profiling offers an aid in 'reducing complexity'[30] for decision-making purposes, this activity can also be a harbinger of risks. In particular, summary decisions or decisions that do not correspond to the actual conditions of the data subject could result from placing the person in a category on the basis of predetermined criteria. Or, again, such placement could generate forms of discrimination or exclusion from access to goods and services[31] . The framework outlined becomes more and more complex when predetermination criteria are established by certain forms of artificial intelligence. In that case, the object of protection becomes the personal identity of the persons concerned and, therefore, their autonomy and freedom of self-determination. In fact, the link between profiling - understood as automated processing capable of defining a subject's 'profile' - and the creation and control that the subject has of his or her personal identity, which increasingly coincides with his or her digital identity,[32] , proves to be very close. With its regulatory and cultural evolution, the protection of personal data has moved, therefore, from an individual dimension to a collective and social dimension: the objective is no longer the mere protection of the individual but also the protection of the community against manipulation and influence by subjects who hold or may potentially hold positions of *particular advantage* and concentration of power[33] . In particular, profiling that is based on predetermined choices may lead to a process of 'normalisation' of the person: in other words, preferences become *sclerotized* precisely on the basis of previous choices revealed by the person concerned himself, so that it even becomes difficult to distinguish which desires are actually proper to the human person and which are instead the product of profiling technologies[34] . Predictive algorithms, together with the artificial intelligences that employ them, can have the most varied uses, including commercial ones. On this point, the doctrine groups such systems into four categories, based on purpose. In particular, there are algorithms directed at activities of

[30] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp.30-31

[31] Pierucci, A. (2019). Data processing and profiling of individuals. Personal data in European law; D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Italian scientific editions.

[32] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p. 31; SARTI, O. S. (2017). Profiling and personal data processing. Technological innovation and the value of the person-The right to the protection of personal data in the EU Regulation 2016, 679, 573, pp.574 ff.

[33] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.31.

[34] Pierucci, A. (2019). Data processing and profiling of individuals. Personal data in European law.

targeted advertising (*targeted advertising*); of *price* discrimination (*price discrimination*); of segmentation or insertion of users into specific groups (*clusters*); of eligibility analysis to obtain certain products (such as, for example, financial and insurance products). In this context, first-, second- and third-degree discrimination can acquire decisive legal significance precisely when the organisational and cataloguing activity of profiling, together with the use of various forms of AI, can give rise to discrimination on both a legal and social level. Therefore, the activity of profiling, in addition to entailing undoubted and numerous advantages (see *above*), can also be a harbinger of criticality and possible discrimination in various spheres: by way of example, think of communications and commercial exchanges on digital platforms; the banking, financial, and insurance sectors; access to Public Administration services; justice; freedom of expression; political propaganda; the formulation of innovative ideas and products; and the maintenance of democratic systems[35] . Especially in relation to the aforementioned risks and critical issues, it appears of primary importance to read Article 4(4) of the GDPR with Article 22 of the Regulation, which refers to so-called automated decision-making processes. Profiling, in particular, stands in relation to them, although not exhausting them, as the most important form of automated processing, which is why it is the only one to be mentioned in Article 22 GDPR .[36]

2.Data protection legislation.

Regulation (EU) No 679 of 2016, adopted by the European Parliament and the Council, was promulgated on 27 April 2016 and published in the Official Journal of the European Union on 4 May 2016. This regulation, which has the title "protection of natural persons with regard to the processing of personal data and the free movement of such data", replaces Directive 95/46/EC, which had been the EU's main legislation on the protection of personal data for several years. The regulation entered into force on 25 May 2018. It is preceded by Recitals, which explain the *rationes* that moved the legislator to legislate on the point, offering a radically changed perspective compared to the previous *status quo*. A neuralgic and delicate point in the legislation at hand is Recital 9, where it is recognised that Directive 95/46/EC

[35] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp. 31-32.

[36] Regulation (EU) 2016/(,79 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation) , Article 22 of which is headed: 'Automated decision-making relating to natural persons, including profiling'; D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.33.

has not succeeded in preventing the fragmentation of personal data protection in the EU, nor has it eliminated legal uncertainty or the perception of risks associated with online transactions[37] . Differences in the implementation of the directive in different Member States have led to different levels of data protection, hampering the free movement of data and restricting economic activities at EU level. In addition to solving these problems, the new regulation responds to the continuous evolution of privacy and data protection concepts, mainly due to the rapid development of online services. The 1995 directive was adopted to protect personal data and ensure their free movement between member states. In the following years, the collection and sharing of data has increased exponentially. Modern technologies now allow operators to use large amounts of personal data for purposes that were unimaginable years ago. In fact, the Internet is now an advertising-supported service, largely based on user profiling. Many users voluntarily publish personal information online, a practice that has become commonplace and has changed social relations, especially for digital natives. This has increased the risk that the online information system can be used for increasingly sophisticated forms of social control. These developments have demonstrated the inadequacy of the 1995 directive, calling for a more robust and coherent legal framework adapted to the development of the digital economy in the internal market. The new regulation applies to any form of processing, whether automated, partially automated or non-automated, of personal data contained in a filing system, as specifically regulated in Article 2 of the regulation[38] . Only the exclusions specified in Art. 2 are exempt, which include

[37] The text of Recital 9) of the GDPR reads 'Although its objectives and principles remain valid today, Directive 95/46/EC has not prevented the fragmentation of the application of personal data protection across the territory of the Union, nor has it eliminated legal uncertainty or the widely held public perception that online operations in particular pose risks to the protection of individuals. The coexistence of different levels of protection of the rights and freedoms of natural persons, in particular the right to protection of personal data, with regard to the processing of such data in the Member States may hinder the free flow of personal data within the Union. Such differences may therefore constitute an obstacle to the exercise of economic activities on a Union scale, distort competition and prevent national authorities from fulfilling their obligations under Union law. This gap in levels of protection is due to differences in the implementation and application of Directive 95/46/EC."

[38] Article 2 of EU Regulation 2016/679 reads: "1. This Regulation shall apply to the processing of personal data wholly or partly by automatic means and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system.

2. This Regulation shall not apply to the processing of personal data:

a) carried out for activities which fall outside the scope of Union law

b) carried out by Member States in the course of activities which fall within the scope of Chapter 2 of Title V TEU

c) carried out by a natural person in the exercise of activities of an exclusively personal or household nature

d) carried out by competent authorities for the purpose of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and prevention of threats to public security.

3. For the processing of personal data by Union institutions, bodies, offices and agencies, Regulation (EC) No 45/2001 shall apply. Regulation (EC) No 45/2001 and other legal acts of the Union applicable to such

processing carried out by Member States for activities related to the common foreign and security policy, or by competent authorities for the prevention, investigation, detection or prosecution of criminal offences, or by a natural person for purely personal or domestic activities (so-called *household exclusion provision*). Article 4 of the regulation[39] provides the

processing of personal data shall be adapted to the principles and rules of this Regulation in accordance with Article 98.

4. This Regulation is therefore without prejudice to the application of Directive 2000/31/EC, in particular the rules on the liability of intermediary service providers set out in Articles 12 to 15 of that Directive."

[39] Thus reads Article 4 GDPR : "For the purposes of this Regulation:

1)"personal data" shall mean any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, by reference in particular to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to his or her physical, physiological, genetic, mental, economic, cultural or social identity;

2)'processing' means any operation or set of operations which is performed upon personal data or sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

3)'restriction of processing' means the marking of stored personal data with the aim of limiting their processing in the future;

4)'profiling' means any form of automated processing of personal data consisting of the use of such personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects of that natural person's professional performance, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements

5)'pseudonymisation' means the processing of personal data in such a way that personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and subject to technical and organisational measures to ensure that such personal data is not attributed to an identified or identifiable natural person

6)"archive' means any structured set of personal data which are accessible according to specified criteria, regardless of whether such set is centralised, decentralised or dispersed on a functional or geographical basis

7)"controller' means the natural or legal person, public authority, agency or other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria applicable to his designation may be established by Union or Member State law

8)'controller' means the natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

9)"recipient' shall mean the natural or legal person, public authority, agency or other body receiving personal data, whether a third party or not. However, public authorities which may receive communication of personal data in the context of a specific investigation in accordance with Union or Member State law are not considered recipients; the processing of such data by those public authorities is in accordance with the applicable data protection rules according to the purposes of the processing

10)'third party' means any natural or legal person, public authority, agency or other body other than the data subject, the controller, the processor and the persons authorised to process personal data under the direct authority of the controller or the processor

11)"Consent of the data subject' shall mean any freely given, specific, informed and unambiguous indication of the data subject's wishes, whereby the data subject, by a statement or an unambiguous affirmative action, indicates his/her agreement to personal data relating to him/her being processed;

12)"personal data breach' means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed

13)"genetic data' means personal data relating to hereditary or acquired genetic characteristics of a natural person which provide unambiguous information on the physiology or health of that natural person, and which result in particular from the analysis of a biological sample of that natural person

14)'biometric data' means personal data obtained by specific technical processing relating to physical, physiological or behavioural characteristics of a natural person which enable or confirm their unambiguous identification, such as facial image or dactyloscopic data

15)"(15) 'health-related data' means personal data relating to the physical or mental health of a natural person, including the provision of health care services, which reveal information relating to the health status of that person

essential definitions of 'personal data' and 'processing'. Personal data means any information concerning an identified or identifiable natural person (the so-called data subject). Particular attention is paid to specific categories of personal data, such as 'genetic data', relating to the genetic characteristics of a natural person; 'biometric data', which allow the unambiguous identification of a person through physical, physiological or behavioural characteristics; and 'health data', whether physical or mental. These categories of personal data are governed by

16) "main establishment" means
   a) in relation to a controller with establishments in more than one Member State, the place of its central administration in the Union, unless decisions on the purposes and means of the processing of personal data are taken in another establishment of the controller in the Union and the latter establishment has the power to order the implementation of those decisions, in which case the establishment which has taken such decisions shall be considered to be the main establishment
   b) in relation to a controller with establishments in more than one Member State, the place where its central administration in the Union is located or, where the controller does not have a central administration in the Union, the establishment of the controller in the Union where the main processing activities are carried out in the context of the activities of an establishment of the controller in so far as that controller is subject to specific obligations under this Regulation;
17) "representative' means the natural or legal person established in the Union who, designated by the controller or processor in writing pursuant to Article 27, represents them in relation to their respective obligations under this Regulation
18) "undertaking' means any natural or legal person, regardless of its legal form, engaged in an economic activity, including partnerships or associations regularly engaged in an economic activity
19) "enterprise group' means a group consisting of a parent undertaking and undertakings controlled by it;
20) "binding corporate rules' means the personal data protection policies applied by a controller or processor established on the territory of a Member State to the transfer or set of transfers of personal data to a controller or processor in one or more third countries, in the context of a business group or a group of undertakings carrying on a common economic activity
21) "supervisory authority' means the independent public authority established by a Member State in accordance with Article 51;
22) "supervisory authority concerned" means a supervisory authority which is affected by the processing of personal data because
   a) the controller or processor is established on the territory of the Member State of that supervisory authority
   b) data subjects residing in the Member State of the supervisory authority are or are likely to be substantially affected by the processing; or
   c) a complaint has been lodged with that supervisory authority;
23) "cross-border processing" means
   a) processing of personal data which takes place in the course of the activities of establishments in more than one Member State of a controller or processor in the Union where the controller or processor is established in more than one Member State; or
   b) processing of personal data which takes place in the context of the activities of a single establishment of a controller or processor in the Union, but which affects or is likely to affect substantially data subjects in more than one Member State;
24) "relevant and reasoned objection' means an objection to the draft decision as to whether or not there is an infringement of this Regulation, or whether or not the action envisaged in relation to the controller or processor complies with this Regulation, which objection clearly demonstrates the significance of the risks posed by the draft decision with regard to the fundamental rights and freedoms of data subjects and, where applicable, the free movement of personal data within the Union
25) "information society service" means a service as defined in Article 1(1)(b) of Directive (EU) 2015/1535 of the European Parliament and of the Council [19]
26) "international organisation" means an organisation and bodies governed by public international law subordinate to it or any other body established by or on the basis of an agreement between two or more States."

Articles 9 and 10 GDPR[40] , which generally prohibit the processing of data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and include genetic data, biometric data intended to uniquely identify a person, data concerning health, sex life or sexual orientation. Exceptions to these prohibitions are specified in Art. 9(2) and include situations where the processing is necessary to protect a vital interest of the data subject or of another natural person, where the data subject is unable to give consent, or where the processing is carried out by foundations, associations or other non-profit bodies with a political, philosophical, religious or trade-union aim, and concerns only the members of those bodies, without communication of the data to the outside world without the data subject's consent. Processing is defined as any operation or set of operations which is performed upon personal data or sets of personal data, whether or not by automated means. This includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data. Getting to the heart of the new legislation, it is important to examine the fundamental principles underpinning it, set out in Chapter II of the regulation, i.e. the provisions set out in Articles 5-11. Article 5[41] , in particular, lists the principles applicable to processing through a set of criteria concerning personal data. According to the first paragraph, personal data must be processed lawfully, fairly and transparently towards the data subject; collected for specified, explicit and legitimate purposes; and not processed in a

[40] Articles 9 and 10 of EU Regulation 2016/679.

[41] The provision in Article 5 GDPR states, "1. Personal data shall be:

a) processed lawfully, fairly and transparently towards the data subject ("lawful, fair and transparent");

b) collected for specified, explicit and legitimate purposes, and further processed in a way that is not incompatible with those purposes; further processing of personal data for archiving in the public interest, scientific or historical research or statistical purposes shall not, in accordance with Article 89(1), be considered incompatible with the original purposes ('purpose limitation')

c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ("data minimisation")

d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ("accuracy")

e) kept in a form which permits identification of data subjects for no longer than it is necessary for the purposes for which the data are processed; personal data may be kept for longer periods provided that they are processed solely for archiving in the public interest, scientific or historical research or statistical purposes in accordance with Article 89(1), subject to the implementation of appropriate technical and organisational measures required by this Regulation to protect the rights and freedoms of the data subject ("conservation limitation");

f) processed in a manner that ensures appropriate security of personal data, including protection, by appropriate technical and organisational measures, against unauthorised or unlawful processing and accidental loss, destruction or damage ('integrity and confidentiality').

2. The controller shall be responsible for compliance with paragraph 1 and able to provide evidence thereof ('accountability')."

way incompatible with those purposes. In addition, data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes of processing; they must be adequate, relevant and limited to what is necessary for the purposes of processing; accurate and kept up to date, when necessary; and processed in a way which ensures appropriate security, including protection against unauthorised or unlawful processing, accidental loss, destruction or damage. The second paragraph of the same article introduces a further principle, stating that 'the controller is responsible for compliance with paragraph 1 and must be able to prove it'. This principle, defined as 'responsibility' in the Italian version of the regulation, is rendered as 'accountability' in the English version. In order to isolate the most novel elements of the new legislation, it is useful to focus on a few fundamental principles: the lawfulness of processing, transparency in processing, the right to be forgotten, the responsibility of the data controller (accountability), and the principles of *privacy by design* and *privacy by default*. These principles are particularly relevant from a privacy point of view, dealt with mainly in Chapter III (Articles 12-23), devoted to the rights of the data subject. The principle of lawfulness of data processing is detailed in Article 6 of the regulation[42] . This article states that the processing of personal data is lawful only if it is based

[42] This is the text of Article 6 GDPR: "1. Processing shall be lawful only if and to the extent that at least one of the following conditions is met:

a) the data subject has given his or her consent to the processing of his or her personal data for one or more specific purposes;

b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or

c) processing is necessary for compliance with a legal obligation to which the data controller is subject;

d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;

e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

f) processing is necessary for the purposes of carrying out a task carried out in the public interest or in the exercise of official authority vested in the data controller; f) processing is necessary for the purposes of the legitimate interests pursued by the data controller or by a third party, provided that the interests or the fundamental rights and freedoms of the data subject requiring the protection of personal data are not overridden, in particular where the data subject is a child.

Point (f) of the first subparagraph shall not apply to the processing of data carried out by public authorities in the performance of their tasks.

2. Member States may maintain or introduce more specific provisions to adapt the application of the rules of this Regulation with regard to the processing, in accordance with points (c) and (e) of paragraph 1, by determining more precisely specific requirements for the processing and other measures to ensure lawful and fair processing also for the other specific processing situations referred to in Chapter IX.

3. The basis for the processing of data referred to in paragraph 1(c) and (e) must be established:

a) by Union law; or

b) by the law of the Member State to which the controller is subject.

The purpose of the processing shall be determined in that legal basis or, as regards the processing referred to in paragraph 1(e), shall be necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. This legal basis could contain specific provisions to adapt the application of the rules of this Regulation, including: the general conditions relating to the lawfulness of the processing by the controller; the types of data to be processed; the data subjects; the entities to whom personal data may be disclosed and the purposes for which they are disclosed; purpose limitations, storage periods and processing operations and procedures, including measures to ensure lawful and fair processing, such as those

on one of two alternative requirements: the necessity of the processing or the consent of the data subject. In other words, processing is only allowed if it is necessary for certain purposes or if the data subject has given his or her specific consent. The cases in which processing is considered necessary are specified by Article 6 itself. For example, processing is necessary for the performance of a contract to which the data subject is party, for the safeguarding of the vital interests of the data subject or of another natural person (as in medical emergencies), or to comply with a legal obligation to which the data controller is subject (as in the case of a notary who has to process the personal data of clients for the advertising of a notarial deed). Art. 4[43] of the regulation defines consent as a free, specific, informed and unambiguous manifestation of will by the data subject, whereby the data subject expresses his or her assent to the processing of his or her personal data by means of a statement or an unambiguous affirmative action. Article 7 regulates both the phase preceding consent (pre-consent phase) and the phase following consent (phase of possible withdrawal of consent). Regarding the pre-consensual stage, the regulation states that if consent is given in the context of a written declaration that also covers other matters, the request for consent must be clearly distinguishable from other matters, understandable, easily accessible and written in plain and clear language. If the declaration violates the regulation, it is not binding, rendering the consent ineffective. Furthermore, Article 7[44] provides that the data subject must be informed

for other specific processing situations referred to in Chapter IX. Union or Member State law pursues an objective of public interest and is proportionate to the legitimate objective pursued.

4. Where processing for a purpose other than that for which the personal data were collected is not based on the consent of the data subject or on a legislative act of the Union or of the Member States which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), in order to verify whether processing for another purpose is compatible with the purpose for which the personal data were initially collected, the controller shall take into account, inter alia

a) any link between the purposes for which the personal data were collected and the purposes of the further processing envisaged

b) the context in which the personal data were collected, in particular in relation to the relationship between the data subject and the controller

c) the nature of the personal data, especially if special categories of personal data within the meaning of Article 9 are processed, or if data relating to criminal convictions and offences within the meaning of Article 10 are processed

d) the possible consequences of the further processing envisaged for the data subject

e) the existence of appropriate safeguards, which may include encryption or pseudonymisation."

[43] Article 4 of EU Regulation 2016/679.

[44] The text of Article 7 of EU Regulation 2016/679 reads as follows : "1. Where the processing is based on consent, the controller must be able to demonstrate that the data subject has given consent to the processing of his or her personal data.

2. Where the data subject's consent is given in the context of a written statement which also relates to other matters, the request for consent shall be presented in a manner clearly distinguishable from the other matters, in a comprehensible and easily accessible form, using plain and intelligible language. No part of such a statement that constitutes an infringement of this Regulation shall be binding.

3. The data subject shall have the right to withdraw consent at any time. Withdrawal of consent shall not affect the lawfulness of processing based on consent before withdrawal. Before giving consent, the data subject shall be informed thereof. Consent shall be withdrawn as easily as it is given.

of the possibility of withdrawing consent at any time, without prejudice to the lawfulness of the processing based on consent prior to withdrawal. Withdrawing consent must be as simple as granting it, without any additional burden or formality. Finally, from an evidential point of view, Article 7 states that the data controller must be able to prove that the data subject has consented to the processing of his or her personal data, implementing the principle of *accountability*. While the regulation provides a detailed regulation for both the pre-consensual and post-consensual phases, consent itself, as a free act of will, is only defined in Art. 4 and not fully regulated. In this respect, the only relevant provision is the interpretative rule in Article 7(4), which states that, when assessing freedom of consent, one must consider whether the performance of a contract or the provision of a service is conditional on consent to the processing of personal data not necessary for such performance. This rule applies in cases where the user is faced with an obligatory choice: consent cannot be withheld in order to use the service. Moreover, consent must relate to data processing that is not strictly necessary for the provision of the service. Situations of this kind are common, for example, when access to an online service is subject to consent not only to the processing of data by the service provider, but also to their transmission to third parties for purposes such as sending commercial communications. The European legislator has provided that if consent is given in such circumstances, it is reasonable to doubt that it was freely given. This is not a presumption, but an invitation to carefully consider this form of conditioning as a possible indication of reduced freedom of decision of the user. This approach is supportable because it rejects the presumption that the data subject's consent is necessarily not free merely because it is considered a requirement for access to a service. Freedom of consent can only be excluded if the service is indispensable for the user or if the provider is obliged to contract. Conversely, freedom of consent is implied when there is a clear instrumental connection between the data processing and the service offered. If this connection is strong, the condition of consent on the part of the service provider is legitimate. An example of this link is found in social networks, whose activities include profiling users' data and preferences in order to organise personalised promotional content. This approach is similar to that adopted by the Garante per la protezione dei dati personali, according to which consent is free when it is not pre-set and is not compulsory in order to benefit from the service offered by the data controller. Indeed, consent is not free if it is necessary for further processing of personal

4. In assessing whether consent has been freely given, the utmost account shall be taken of the possibility, inter alia, that the performance of a contract, including the provision of a service, may be affected by the provision of consent to the processing of personal data not necessary for the performance of that contract."

data as a condition for obtaining a requested service. The issue of free consent becomes even more delicate in relation to profiling, which consists of any form of automated processing of personal data to evaluate personal aspects of a natural person. This includes the analysis or prediction of aspects such as professional performance, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements. The regulation severely restricts this particularly widespread and economically relevant practice. Article 22 states that the data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects or significantly affects him or her. This rule is tempered by some exceptions: the explicit consent of the data subject, legal authorisation, or the necessity of the decision for the conclusion or performance of a contract between the data subject and the controller. Again with regard to the consent of the data subject, it is important to stress that the legislator, in Article 8 of the regulation[45] , has paid special attention to the consent of minors with regard to information society services. If the processing of children's data is based on their consent, this is only considered lawful if the child is at least 16 years old. For children under the age of 16, the processing is lawful only if it is given or authorised by the holder of parental responsibility. Member States may, however, set a lower age by law, provided that it is not less than 13 years old. The regulation specifies that the child consent rule is without prejudice to general contract law provisions of the Member States, such as the rules on the validity, formation or effect of a contract in relation to a child (Art. 8(3)). This clarification is appropriate in order to keep the two areas separate - the contractual relationship between the parties and the relationship between the data subject and the controller. The contractual relationship follows the rules of the negotiated agreement and moves along the lines of the relevant regulation, whereas the relationship between the data subject and the data controller is based on the consent of the former and follows rules of legal origin and of an essentially imperative nature. The principle of transparency is at the heart of Articles 12 et seq. concerning the relationship between the data controller and the data subject, with specific reference to information and access to data. The implementation of this principle is mainly entrusted to the prescription of specific

[45] The text of Article 8 GDPR reads, "1. Where Article 6(1)(a) applies in respect of the direct offering of information society services to children, the processing of personal data of a child shall be lawful where the child is at least 16 years of age. Where the child is below 16 years of age, such processing shall be lawful only if and to the extent that such consent is given or authorised by the holder of parental responsibility.
Member States may lay down by law a lower age for these purposes provided that it is not less than 13 years.
2. The controller shall take all reasonable steps to verify in such cases that consent is given or authorised by the holder of parental responsibility over the child, having regard to the available technology.
3. Paragraph 1 shall be without prejudice to the general provisions of the contract law of the Member States, such as the rules on the validity, formation or enforceability of a contract in relation to a child."

information modalities and a detailed proceduralisation thereof. Information intended for the public or the data subject must be easily accessible and comprehensible, expressed in simple and clear language. This is particularly relevant as regards informing data subjects of the identity of the data controller and the purposes of the processing, as well as further information concerning the data subject's right to obtain confirmation and communication of the processing of personal data concerning him or her. These provisions are part of the broader tendency of the Community legislator to promote user information by imposing clear communication obligations on operators. In this specific area, this aims to ensure that users are on the one hand made aware of the risks and informed of the guarantees relating to the processing of their data, and on the other hand, informed about the specific purposes of the processing, enabling them to assess whether the data requested and the way in which they are collected and stored are actually in line with those purposes. Interestingly, the principle of transparency is the normative basis for the use of so-called multi-layered formats. As is well known, *privacy policies* are complex documents containing a large amount of information of a technical-legal nature. Multilayer notices aim to improve the quality of information by condensing everything the data subject needs to understand his position and make decisions into simplified information 'layers': in this way, he should be able to understand how his data are used at a glance. This is provided for in Article 127[46] , which states that 'the information to be provided to data subjects (...) may be provided in combination with standardised icons in order to give, in an easily visible, intelligible and clearly legible manner, an overview of the intended processing. If presented electronically, the icons shall be machine-readable'. Finally, it is worth mentioning the issue, closely linked to that of transparency, of certification mechanisms, seals and marks of data protection, aimed at enabling data subjects to quickly assess the level of protection of their data: these are regulated in Chapter IV of the regulation, specifically in Article 42[47] . The right to be

[46] Art. 127 EU Regulation 2016/679 .

[47] Art. 42 GDPR provides: "Member States, supervisory authorities, the Committee and the Commission shall encourage, in particular at Union level, the establishment of data protection certification mechanisms as well as data protection seals and marks for the purpose of demonstrating the compliance of processing operations carried out by data controllers and processors with this Regulation. The specific needs of micro, small and medium-sized enterprises shall be taken into account.

2. In addition to the adherence of the controllers or processors subject to this Regulation, mechanisms, seals or marks approved in accordance with paragraph 5 of this Article may be put in place to demonstrate the provision of appropriate safeguards by controllers or processors not subject to this Regulation pursuant to Article 3, in the context of transfers of personal data to third countries or international organisations under the conditions set out in Article 46(2)(f). Such controllers or processors shall give a binding and enforceable undertaking, by contractual or other legally binding means, to apply the same appropriate safeguards also with regard to the rights of data subjects.

3. Certification shall be voluntary and accessible through a transparent procedure.

forgotten, in the specific context under consideration, represents the data subject's right to the rectification and erasure of personal data relating to him or her. This right is considered absolute and can only be limited in the presence of interests specifically identified by the legislator. As regards the theoretical content of the right to be forgotten, doctrine has formulated various hypotheses. In some respects, this may be considered a natural consequence of the correct application of the general principles of the right to report news: just as one should not divulge a fact without real public interest, in the same way one should not republish an old piece of news that is no longer relevant to current information. In other respects, the right to be forgotten could be seen as an aspect of the more general right to privacy or as a profile of the right to honour. In short, the common feature of all these interpretations is the protection of the subject's interest in not preserving a collective memory of outdated or falsely attributed events. The regulation, however, seems to outline a broader concept of the right to be forgotten, considering it as the general right to have one's personal data deleted in various circumstances. Article 17 of the regulation[48] states that 'the data subject shall have the right to obtain from the controller the erasure of personal data concerning him without undue delay and the controller shall be obliged to erase the personal data without undue delay'. The circumstances constituting the prerequisite for the right to be forgotten include situations of both an objective and subjective nature: among the former, there is the non-necessity of the data in relation to the purposes for which they were collected, unlawful processing of the data and the legal obligation to erase; among the latter, there is the data subject's withdrawal of consent without any other legal basis for processing, and the data subject's objection to processing. However, the right to be forgotten cannot be exercised when the data processing is necessary for the exercise of the right to freedom of

4. Certification pursuant to this Article shall not reduce the responsibility of the controller or processor for compliance with this Regulation and shall be without prejudice to the tasks and powers of the competent supervisory authorities pursuant to Articles 55 or 56.
5. Certification under this Article shall be issued by the certification bodies referred to in Article 43 or by the competent supervisory authority on the basis of criteria approved by that competent supervisory authority in accordance with Article 58(3) or by the Committee in accordance with Article 63. Where the criteria are approved by the Committee, this may result in a common certification, the European Data Protection Seal.
6. The controller or processor submitting the processing operation to the certification mechanism shall provide the certification body referred to in Article 43 or, where applicable, the competent supervisory authority with all information and access to the processing activities necessary to carry out the certification procedure.
7. The certification shall be granted to the controller or processor for a maximum period of three years and may be renewed under the same conditions provided that the relevant requirements continue to be met. Certification shall be withdrawn, where applicable, by the certification bodies referred to in Article 43 or by the competent supervisory authority, as appropriate, if the requirements for certification are not or are no longer fulfilled.
8. The Board shall collect all certification mechanisms and data protection seals and marks in a register and make them public by any appropriate means."

[48] Art. 17 GDPR

expression and information, for the fulfilment of a legal obligation, for reasons of public interest in the field of public health, for archiving in the public interest, for research or statistical purposes, or for the establishment, exercise or defence of a legal claim. It is complex to bring the prerequisites of the right to be forgotten identified by the regulation into a unity *of rationale* and, consequently, to reconstruct this right as a whole. Although it can be partly traced back to the profiles traditionally evoked by doctrine, it is also conceived by the Community legislator as a defence of the data subject against the unlawfulness of processing or its mere uselessness. The implementation of the accountability principle is described in Chapter IV of the regulation, devoted to the controller and the processor. In particular, the latter is defined by Article 4 as the person who 'processes personal data on behalf of the controller'. Article 24[49] , entitled 'Responsibilities of the controller', states that the controller must take appropriate technical and organisational measures to ensure and demonstrate that processing is in compliance with the regulation. These measures must be reviewed and updated when necessary. The concept of accountability is difficult to translate into a single Italian word and is rendered in the Italian version of the regulation as 'responsibility'. However, it lies somewhere between accountability and compliance, since the holder must be compliant with the applicable legislation. With regard to the protection of personal rights, this principle may seem to be a corollary of the principle of transparency. However, accountability must be understood not as complete accessibility of information for the data subject, but as a guarantee of the compliance of the activity with the sectoral regulation, a compliance that the data controller must be able to prove at any time. From this point of view, the adherence of the data controller to a code of conduct offers a facilitation: Article 243 of Regulation[50] states that 'adherence to codes of conduct [...] or to a certification mechanism [...] may be used as an element to demonstrate compliance with the obligations of the data controller'. Beyond the evidentiary level, it seems that the accountability principle requires data controllers to adopt a data protection control system, structured according to universally recognised standards of good administration and externally verifiable, since compliance with the rules on personal data requires not only mere compliance with legal

[49] Art. 24 of EU Regulation 2016/679 reads as follows: "1. Having regard to the nature, scope, context and purposes of the processing, and having regard to the risks having different probability and severity for the rights and freedoms of natural persons, the controller shall implement appropriate technical and organisational measures to ensure, and be able to demonstrate, that the processing is carried out in accordance with this Regulation. Those measures shall be reviewed and updated when necessary.
2. Where proportionate in relation to the processing activities, the measures referred to in paragraph 1 shall include the implementation of appropriate data protection policies by the controller.
3. Adherence to the codes of conduct referred to in Article 40 or to a certification mechanism referred to in Article 42 may be used as an element to demonstrate compliance with the obligations of the controller."
[50] Art 243 GDPR.

provisions, but also the setting up of adequate internal governance. Codes of conduct are regulated in Section 5 of Chapter IV of the regulation (Art. 40-43), which is also dedicated to the certification of data controllers. In particular, Article 40[51] requires Member States to encourage the drawing up of such codes - or the adherence to already existing codes - by associations and other bodies representing categories of data controllers; in addition, Member States are also required to encourage the setting up of data protection certification mechanisms, as well as of data protection seals and marks to demonstrate the compliance of processing operations carried out by data controllers and data processors with the regulation, as laid down in Article 41[52] . To this end, Article 43[53] regulates the appropriate certification bodies. Fundamental to the economy of the regulation is the principle of *privacy by design*, which states that data protection must be integrated into the entire life cycle of a technology, service or process, starting from the design phase. This implies that every project must be designed from the outset with the privacy of the end user and the protection of his or her personal data in mind, with all necessary supporting applications, both IT and non IT. This approach aims to ensure maximum effectiveness in data protection and is increasingly adopted in the field of personal information management. A complementary but distinct principle is that of *privacy by default*. This principle requires that data be collected to the minimum necessary extent and that the purposes of processing be as limited as possible. In essence, this involves applying the principles of data minimisation and purpose limitation of processing. Also derived from these principles is storage limitation, which requires that the period of processing and storage of collected data be kept to a minimum. Both principles, *privacy by design* and *privacy by default*, have been incorporated into the European Data Protection Regulation (GDPR). The recitals of the regulation state that the data controller should adopt internal policies and measures that meet the principles of data protection by design and privacy by default. Such measures may include, inter alia, minimising the processing of personal data, pseudonymising data as soon as possible and improving security features during the development, design and use of applications, services and products based on data processing. Manufacturers are encouraged to take the right to data protection into account when developing and designing these products, services and applications. Article 25 of the GDPR[54] forms the regulatory core of these principles. It states that the data controller must take appropriate technical and organisational measures, such as pseudonymisation, to

[51] Art 40 GDPR.
[52] Art 41 GDPR.
[53] Art 43 GDPR.
[54] Art 25 GDPR.

effectively implement the data protection principles, including minimisation, and integrate the necessary safeguards into the processing to meet the requirements of the Regulation and protect the rights of data subjects. In addition, the controller must ensure that, by default, only the personal data necessary for each specific purpose of processing are processed. This obligation concerns the amount of personal data collected, the scope of the processing, the storage period and accessibility. In order to demonstrate compliance with these principles, the controller may make use of certification mechanisms provided for in Article 42 of the regulation (see *above*). Chapter VIII of the regulation, which includes Articles 77 to 84, is devoted to remedies, liability and sanctions. In this context, Article 82[55] , which mainly refers to the right to compensation and liability, lays down the general principle that any person who suffers material or immaterial damage as a result of a breach of the regulation has the right to obtain compensation from the controller or processor. The controller is only liable for the damage caused if he has not fulfilled the obligations specifically addressed to him or has acted in a manner inconsistent with or contrary to the lawful instructions of the controller. The regulation also provides for a system of joint and several liability in the case of concurrent liability. It is stipulated that if several controllers or processors are involved in the same processing and responsible for the damage, each of them shall be jointly and severally liable for the entire amount of the damage. This regime aims to ensure the effective compensation of the data subject. If one of the data controllers or persons responsible has paid full compensation, he or she is entitled to claim from the others involved the part corresponding to their liability, determined according to Article 2055 of the Civil Code[56] . The regulation also contemplates the possibility of a liberating proof, exonerating both the owner and the liable party from liability if they prove that the harmful event was not their fault. This assertion derives from the fact that Article 82 may be interpreted in at least two ways in order to coordinate it with the categories of the Italian civil liability system: a first reading suggests that the Community legislator, in dealing with imputability, refers to the criterion of Article 2046 of the Civil Code, or to the causal link between the conduct of the owner or the person liable and the damage. This interpretation, however, would empty the regulation's provision of content, rendering it a pleonasm. A second reading, on the other hand, might hold that the exemption depends on the active conduct of the agent, consisting

[55] Art 82 GDPR.

[56] Art. 2055 of the Civil Code states: 'If several persons are jointly and severally liable for the damage, they shall all be jointly and severally liable to pay compensation.
The one who has compensated the damage shall have recourse against each of the others, to the extent determined by the seriousness of their respective fault and the extent of the consequences resulting therefrom. In case of doubt, the individual faults shall be presumed to be equal."

of having done everything possible to avoid the damage, analogous to what is provided for in various hypotheses of strict liability in the Italian legal system (e.g. Articles 2047 and 2048 of the Civil Code). Ultimately, the regulation is both a starting point and a significant challenge. Its main objective is to ensure a uniform and high level of personal data protection for European citizens in all EU Member States, thereby increasing trust in public administrations and private economic operators with regard to data protection. The regulation required economic operators to go beyond the purely formal aspects of implementing the legislation, promoting a profound cultural change. This is necessary for business practices to continually adapt to the changes resulting from inexorable technological progress (e.g. cloud computing, social media and the interconnection of databases). These requirements have inevitably had a significant organisational impact: consider, for example, the adoption of the principle of *privacy by design*, which aims to reshape the design models of services, software and business processes. The ultimate goal, though ambitious, is of great value: to facilitate both the users of services and the economic operators themselves, who will benefit from a single privacy regulation. This process also benefits the market, which continues to be the fulcrum around which much of the regulatory system and EU policies revolve[57] .

2.1. Automated processing between opportunity and ambiguity: the use of predictive algorithms.

Focusing, in particular, on the scope of Articles 4 and 22 of the Regulation, it is clear that the creation of ever more accurate profiles is a strategic activity, especially in sectors that are primarily characterised by the so-called *data-driven economy*[58] . In this context, profiling activity arises as capable of justifying and ordering the enormous flow of data present; it also enables the extraction of information from such data by means of algorithmic operations, and allows artificial intelligence tools to enhance and make use of the aforesaid information. In this way, an inseparable link is established between the preparatory and preliminary activity of profiling and that integrating automated processing that in various ways leads to decisions that may

[57] For the entire paragraph see Ciullo, M. Profiling and automated decisions: theoretical and evolutionary profiles, forthcoming.
[58] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.34, See also Frosini, T. E. (2020). Constitutionalism in the technological society. Jurcost. org, Liber Amicorum per Pasquale Costanzo, 25.

significantly affect the individual . [59]It is clear, on the other hand, that the activity of profiling and that of automated decision-making processes remain distinct; however, it cannot be denied that technological development and market requirements lead to at least a partial and necessary overlap of the two activities. The regulation looks precisely at this nuance: in fact, profiling is considered as automated processing. It is this close correlation and this almost total overlap that makes one discuss the opportunities and ambiguities associated with the processing of personal data[60] , also considering it with reference to the most recent advances in technology. Risks and benefits combine to create a framework of considerable interest from a legal point of view. From the point of view of 'risks', automated processing could be considered a 'dangerous activity' in the civil law sense, or 'present a high risk for the rights and freedoms of natural persons', thus requiring a data protection impact assessment, as reflected in Articles 35 and 36 of the Regulation[61] . As far as the benefits are concerned, it would be anachronistic and unobjective not to claim that data processing, even by means of artificial intelligence tools, does not bring highly efficient services and a general improvement of human life. On the other hand, it is, however, worth emphasising that a dangerous drift of the activity of profiling, as the main form of automated processing, could be a pervasive surveillance and massive manipulation of individuals. In other words, profiling might not be limited to the prediction of behaviour, but could go so far as to actually influence individuals, inducing them to behave or perform actions that they would not otherwise have behaved or performed. This prompted, among other things, deep reflection and analysis on the interaction of these dynamics with the protection and enhancement of fundamental rights[62] . More in depth, it seems clear that the new artificial intelligence tools offer unprecedented opportunities for human beings that, necessarily, pass through the use of profiling, i.e. the possibility of collecting and linking information on individual consumers in order to derive behavioural patterns and facilitate and direct choices[63] . The prediction of future behaviour offered by predictive algorithms is, in any case, based on past behavioural patterns and choices, reworked according to probabilistic techniques: in the presence of inaccurate,

[59] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p. 34.
[60] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.34
[61] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.34,
[62] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.34 and also BUSIA, G. (2013). The frontiers of privacy on the Internet: The new gold rush for personal data. Internet: Rules and protection of fundamental rights.
[63] BUSIA, G. (2013). The frontiers of privacy on the Internet: The new gold rush for personal data. Internet: Rules and the protection of fundamental rights, pp.14 ff.

incorrect or *biased* data processing, it may lead to erroneous predictions or to the crystallisation of a divergent future with respect to the subject referred to[64] . Data analysis and the integration of predictive algorithms are therefore crucial to generate predictions of future behaviour. The process begins with the collection of data, which is then analysed to identify correlations and inferences. The data are processed at different stages, increasing their value. Algorithms, based on typical characteristics and behaviours of individuals, create profiles and categories, improving prediction capability. The accuracy of the predictions depends on the quantity and quality of the data used. This raises legal issues concerning the protection of self-determination, as the use of outdated or inaccurate data could lead to incorrect predictions. It is therefore essential to establish appropriate regulations to address these issues and protect individual rights. This need was taken into account in the wording of Article 22 GDPR, even though the provision - although innovative in scope and appropriate to deal with certain issues - proved insufficient to address certain problems arising from concrete cases[65] . As an example, consider that through the analysis of personal and behavioural data, predictive algorithms can influence both marketing and electoral consent. Algorithms rely on the accumulation of information to create user profiles, thus enabling highly targeted advertising that takes into account searches made and pages visited. This approach, known as *behavioural advertising*, is supported by commercial agreements between various market players. Furthermore, the impact of modern advertising techniques on politics should be emphasised, highlighting the use of sponsorships on social networks and micro-targeting strategies, which can alter electoral consent and raise concerns about voter privacy. Reference can also be made to the risk of mass control through online voting systems, where anonymity would not be guaranteed, highlighting the ethical and legal implications of such practices[66] . Thinking of different and more innovative implementations of artificial intelligence systems, one can think of user data in order to better calibrate the results of certain generative artificial intelligence (see *below*).

3. The evolution of the regulation of profiling and automated decision-making processes.

[64] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p. 36

[65] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p. 38 and Ciullo, M. Profiling and automated decisions: theoretical and evolutionary profiles, forthcoming.

[66] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp. 38 ff.

In a context in which data are not only widely diffused by now, but also act as the fuel of many human activities in digital and integrated dynamics (think of the *Internet of things*), it is more than self-evident to understand how having and governing the algorithm and all its declinations and evolutions represents governing society, individuals with an experience of behaviour, choices and preferences, it represents immediate access to ganglia of institutions, businesses and the very different relations of reality. Hence the need to legislate on the digital environment, *lato sensu* intended, and its constant and rapid evolution. These latter characteristics have made the legislator's activity complex and inherently subject to revisions and reversals. In other words, it is clear that profiling activities and automated decision-making processes can offer efficiency and at the same time benefit various sectors - think of education, health, and transport by way of example - precisely through market segmentation and the personalisation of services. At the same time, the risks for the rights and freedoms of natural persons may increase[67] , at least in cases where there is no obligation of transparency with respect to the above-mentioned phenomena. Indeed, the effect of profiling may be the perpetuation of stereotypes or certain types of classifications that in fact confine a natural person to a certain category and thus limit him or her to his or her preferences, which the digital system appropriates and continues to propagate to him or her. It is clear how this can represent a threat in the freedom of choice not only of products and services and reduce the capacity of self-determination of the natural person in question[68] . Moreover, it is important to emphasise how profiling activity is not limited to intervening merely in *marketing* contexts: as mentioned above, it invests much broader sectors and in which various fundamental rights of the individual come into play[69] . It is no coincidence that *the rationale* that moved the legislator in regulating through Regulation 2016/679 was precisely, among other things, respect for the rights of the subjects involved by striking a balance that always sees economic interests as ancillary to the human person[70] . In other words, ethically, a 'constitutional' value has been assigned to a human-centred Artificial Intelligence and digital ecosystem, whose existence on the level of positive law is often still fractious and unfinished, despite the latest legislations on the subject. The focus can no

[67] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p. 41.
[68] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp 41-42
[69]Poletti, D. (2018). Regulating technology: the EU Reg. 2016/679 and the protection of personal data. A dialogue between Italy and Spain. STUDIES ON THE INTERNET ECOSYSTEM, 1-512, p.30
[70] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.43 et seq. See also Rodotà, S. (1998). Report for the year 1997 of the guarantor for the protection of personal data. information and computing.

longer be placed solely on the General Data Protection Regulation[71] . Even more so is this to be taken into account with reference to further and subsequent technological developments after the entry into force of the aforementioned Regulation. The advent of Artificial Intelligence has in fact represented a test case in the areas of profiling and automated decisions.

3.1. Provisions on profiling and automated processing prior to EU Regulation 2016/679.

The current variegated European legal framework of technological and digital advances is the result of reflections and normative contributions in various senses that have taken place over the previous decades. In general, with regard to the notion of *privacy*, considered in conjunction with the centrality and dignity of the human person, it is necessarily necessary to examine - albeit briefly - the great Charters of Rights that flourished in the second half of the 20th century. Of all of them, it seems appropriate to refer to the 'Universal Declaration of Human Rights' of the General Assembly of the United Nations, proclaimed in Paris on 1 December 1948, which, in its Preamble, recognises dignity as inherent to every human being, as well as the foundation of freedom, justice and peace in the world[72] . It is noteworthy that none of the Constitutions that blossomed in the aftermath of World War II felt the need to include the protection of personal data as a right in its own right, distinct and separate from

[71]Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021), p. VIII.

[72] Thus states the Preamble of the Universal Declaration of Human Rights : "Whereas recognition of the inherent dignity of all members of the human family and of their rights, equal and inalienable, constitutes the foundation of freedom, justice and peace in the world;
Whereas disregard and contempt for human rights have led to barbarous acts that offend the conscience of mankind, and the advent of a world in which human beings enjoy freedom of speech and belief and freedom from fear and want has been proclaimed as man's highest aspiration
Whereas it is imperative that human rights be protected by legal norms if man is not to be forced to resort, as a last resort, to rebellion against tyranny and oppression; and
Whereas it is imperative to promote the development of friendly relations among nations;
Whereas the peoples of the United Nations have in the Charter reaffirmed their faith in fundamental human rights, in the dignity and worth of the human person, in the equal rights of men and women, and have determined to promote social progress and a better standard of living in greater
freedom;
Whereas Member States have pledged to pursue, in cooperation with the United Nations, universal respect for and observance of human rights and fundamental freedoms
Whereas a common understanding of these rights and freedoms is of the utmost importance for the full realization of these commitments;
THE GENERAL ASSEMBLY
proclaim
This Universal Declaration of Human Rights as a common ideal to be attained by all peoples and all nations, to the end that every individual and every organ of society, keeping this Declaration constantly in mind, shall strive by teaching and education to promote respect for these rights and freedoms and by progressive measures, national and international, to secure their universal and effective recognition and respect, both among the peoples of Member States themselves and among the peoples of territories under their jurisdiction."

the protection of privacy and private life[73] ; at the same time, the European Convention on Human Rights, signed in Rome in 1950, is believed to be the first European source to lay the foundations for the protection of personal data. Indeed, traditionally Article 8 of the Charter[74] is regarded as the normative matrix of personal data protection. This protects private and family life, home and correspondence. In fact, Article 8 ECHR relies on the need for a legal basis to justify any interference in the rights it protects. Furthermore, remaining in the international sphere but focusing more specifically on profiling and automated processing, the 'Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data', i.e. Council of Europe Treaty No 108/1981, signed in Strasbourg, and recently updated[75] , is of vital importance with regard to automated processing. This Convention set itself up as the first legal instrument for the protection and safeguarding of personal data: in fact, all subsequent regulatory and jurisprudential interventions with reference to the effects of technological developments on personal data have been based on it[76] . The Treaty we are dealing with here was born in a peculiar context, which saw the rise and spread of tools such as 'databases', capable of storing large numbers of data and combining them with each other in order to be able to exploit the results emerging from the interaction: in other words, through these operations, new information could be obtained from the initial data, new value and usefulness of the data already possessed. Against this background, it is easy to understand how Treaty 108/1981 recognised the correct handling of personal data with reference to automated processing. In addition, the Convention contains rules that refer to issues that will become crucial in the subsequent legislation on the protection of personal data: these include the ban on the processing of so-called 'sensitive data' - later to become 'sensitive data' and 'special categories of data' - , the right to access and rectify one's own data and limitations, and cross-border flows of personal

[73] Pizzetti, F. (2016). Privacy and the European right to personal data protection: From Directive 95/46 to the new European Regulation. G Giappichelli Editore, p.56.

[74] L 'Art. 8 of the European Convention on Human Rights states: "1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right unless such interference is provided for by law and constitutes a measure which, in a democratic society, is necessary for national security, public safety, the economic well-being of the country, for the defence of order and the prevention of crime, for the protection of health and morals, or for the protection of the rights and freedoms of others."

[75]D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.46.

[76] See on this point Formici, G. (2021). La disciplina della data retention tra esigenze securitarie e tutela dei diritti fondamentali: Un'analisi comparata (Vol. 94). Giappichelli, p. 40-41; Sola, A. (2023). Areas of interest for the regulation of data economies in the relationship between law and technology. FEDERALISMS. IT, (10), 195-217 and D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p. 46.

data[77] . Subsequent to the Convention is the so-called 'Mother Directive', i.e. Directive 95/46/EC (see *above*), which continued to deal with automated data processing, albeit without any reference to profiling[78] . In particular, Article 15 of the aforementioned Directive[79] includes among the rights of each individual the right not to be subject to a decision which would have a significant effect on the individual, where the decision is based exclusively on automated processing of data intended to evaluate certain aspects of the personality of the individual[80] . In fact, the *rationale of* this rule is to be found in a basic concern that moved the 1995 legislator: the potential capacity of automated decisions, which are becoming more and more relevant as technology evolves, to reduce the role played by individuals in shaping important decision-making processes that affect them, was considered with particular regard and in a negative sense. The standard, indeed, might have appeared as a compromise attempting to balance the two instances, yet it was perceived as harbouring even more ambiguities[81] . On the basis of this act, two legislative acts were adopted in the Italian legal system: the first, i.e. Law No. 675/1996, regulating the 'Protection of persons and other subjects with regard to the processing of personal data'; the second, i.e. Legislative Decree No. 196/2003, containing the 'Personal Data Protection Code'. Already with the first act, the issue at hand here involved the question of the protection of persons and the proper processing of their personal data. On this point, the doctrine seemed to anticipate the demands that have only recently been defined as *risk scoring*. In particular, it was feared that

[77] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.46.

[78] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.47

[79] L' Art. 15 of Directive 1995/46/EC reads: "1. Member States shall grant to any person the right not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.".

2. Member States shall provide, without prejudice to the other provisions of this Directive, that a person may be subject to a decision as referred to in paragraph 1 if such a decision

(a) is taken in the context of the conclusion or performance of a contract, provided that the request concerning the conclusion or performance of the contract, made by the person concerned, has been satisfied or that appropriate measures, including the possibility to put his point of view, ensure that his legitimate interests are safeguarded; or

(b) it is authorised by a law specifying measures to safeguard a legitimate interest of the person concerned."

[80] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.47 and more extensively Bygrave, L. A. (2001). Automated profiling: Minding the machine: Article 15 of the EC data protection directive and automated profiling. Computer Law & Security Review, 17(1), 17-24.

[81] Bygrave, L. A. (2001). Automated profiling: Minding the machine: Article 15 of the EC data protection directive and automated profiling. Computer Law & Security Review, 17(1), 17-24.

On this point, he offers an even broader reading and a comparison between Article 15 of the so-called Mother Directive and Article 22 GDPR, Wong, B. (2021). Online personalised pricing as prohibited automated decision-making under Article 22 GDPR: a sceptical view. Information & Communications Technology Law, 30(2), 193-207.

the creation of profiles and the reconstruction of personalities by *automated means*, which would lead to decisions implying an assessment of human behaviour, could have a potentially damaging impact to the extent that it could be placed in a 'meta- or super-individual' dimension, thus transcending the sphere of personal identity and the mere protection of privacy. It is, in fact, interesting to note how the doctrine was concerned about drifts that would feed an accentuated conformism, determinism and control, potentially leading to the dystopian creation of 'abstract profiles' that would invest many of the spheres of human living and acting. In other words, they feared the creation of 'ideal models' - of workers, of citizens, of people truly in need of assistance - that could prove to be a harbinger of discrimination with respect to individuals who, in fact and in physical reality, did not correspond to the profiles created. As a result of this law, in the Italian legal system, the right not to be subjected to an assessment based solely on automated processing was subsequently adopted by Legislative Decree No. 196/2003, which made explicit reference to the "definition of profiles and the personality of the person concerned"[82] . In fact, the rule referred to (Article 14 of Legislative Decree 196/2003) regulates automated processing in a manner not dissimilar to what will be the future Article 22 GDPR, containing within it an implicit reference to the concept of profiling[83] . What differs between the two rules is certainly the scope of application: on the one hand, Article 14 of Legislative Decree 196/2003 prohibited automated processing with reference only to the public sphere; on the other hand, Article 22 GDPR - in line with the so-called 'Mother Directive' - applies to any decision that is based solely on automated processing, placing no limitation whatsoever on the nature of the data controller. Consequently, the perspective of framing activities carried out in the digital environment changes radically as the digital environment itself changes. The first paragraph of the aforementioned Article 14 of Legislative Decree No. 196/2003 laid down an 'explicit and absolute' prohibition for the person who took a decision based exclusively on automated processing in order to designate profiles or the personality of the person

[82] Art. 14 Legislative Decree 196/2003 , replacing Art. 17 of Law 675/1996

[83] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp.46-49. In particular, on this point, the Author believes that 'the passage from the Directive to the Regulation has led to a greater awareness of the autonomy of the activity of profiling as one of the possible automated processing, giving it a specific definition and ad hoc discipline for fully automated processing, among which profiling itself could also fall'. The same author goes on to state how the rule of the Privacy Code that he deals with here prohibited automated processing, "or, rather, recognised the data subject's right to oppose it unless the determination
was adopted on the occasion of the conclusion or execution of a contract: or in acceptance of a proposal by the data subject; or on the basis of adequate safeguards identified by the Code or by a provision of the Garante following what was the 'preliminary verification' governed by Article 17. On this point, see Pierucci, A. (2019). Data processing and profiling of individuals. Personal data in European law, p.420

concerned. Paragraph 2 of the same provision provided, specularly, for the data subject's right to object to the processing, with the exception of a decision taken with reference to a contract, accepting a proposal by the data subject or based on guarantees provided for by the Code or by provisions of the Garante. Reading the two paragraphs in conjunction, it was clear that decisions based solely on automated processing in the public sphere were prohibited, while at the same time they were admissible in the private sphere, given the data subject's power to object. In any case, the issue of the repercussions and impacts that automated tools and profiling could produce in the legal sphere of private individuals was already strong in those years, repercussions and impacts that were destined to increase with the evolution of technology and technique[84] . Further references to profiling, mostly understood as a risky practice, can be found in Legislative Decree 196/2003[85] . One may also recall Article 8 of the so-called 'Declaration of Rights on the Internet' of 28 July 2015[86] . It should also be noted that, on the subject of profiling and automated processing, there have been numerous interventions and measures by the National Data Protection Authority[87] . The first decisions on the subject of profiling date back to the early years of the new millennium, but it was in 2005[88] that the Italian Garante per la Protezione dei Dati Personali organically analysed the matter, in connection with the aforementioned legislative decree 19672003. The great work of the Garante anticipated instances and dictated measures for concrete cases as they arose. Subsequently, a few years earlier than the entry into force of EU Regulation 2016/679, the subject of profiling was dealt with in Order No. 161/2015, containing the "Guidelines on the processing of personal data for online profiling"[89] . In that case, the measure relied, among other things, on the 'need to obtain free, informed and always

[84] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp.50-51

[85] D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.51. The regulations referred to by the author are Article 22(10), with reference to the prohibition to process sensitive and judicial data in the context of psycho-aptitude tests aimed at defining the profile or personality of the data subject, and in Article 37(1)(d), in the part which provided for the obligation to 'notify' to the Garante the processing operations carried out with the aid of electronic instruments aimed at defining the profile or personality of the data subject or at analysing habits or consumption choices, or at monitoring the use of electronic communication services. Again, the reference contained in Article 14 of Legislative Decree No. 196/2003 to the preliminary verification referred to in Article 17 of the same code has enabled the Garante d1 over the years to carry out a careful control activity on such processing."

[86] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.51.

[87] On this point see, more extensively, D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp. 52-58. The author offers a careful and exhaustive examination of the contribution of the Guarantor Authority.

[88] GPDP, 'Fidelity cards' and consumer guarantees. The Garante's rules for fidelity programmes, 24 February 2005, web doc. no. 1103045.

[89] GPDP, Guidelines on the processing of personal data for online profiling, 19 March 2015, web doc no. 3881513.

revocable consent'; secondly, it should be noted that that measure specified certain differences, also taking into account their legal relevance, between two profiling systems: *cookies* and *fingerprinting*. Finally, one should consider the 'Mevaluate' project, whereby reputations concerning natural and legal persons were to be entered into an online database by a private platform[90] . With regard to this activity, the *authority* took the view that it was a 'reputational *rating*' activity, and therefore a violation of personal dignity. This position - following a judicial process that lasted several years - was endorsed by the jurisprudence of the Court of Cassation with judgment No. 14381 of 24 May 2021 .[91]

3.2. Profiling activities and automated decision-making in the EU Regulation 2016/679.

As mentioned *above* (see *above*), the disruptive scope of the General Data Protection Regulation, which entered into force in 2018, also found its raison d'être in the systematic and organic definition, for the first time, of profiling activity. The regulation also contains regulations concerning decisions taken through automated processing - first and foremost, profiling - further distinguishing between solely automated processing and only partially automated processing. At the same time, profiling and automated decision-making processes are also regulated by Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data .[92]

Article 4(4) of the Regulation defines profiling as:

"any form of automated processing of personal data consisting of the use of such personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects of that natural person's professional performance, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".

From this it can be deduced that profiling consists of three essential elements: firstly, the processing of personal data; the processing carried out by automated means; and the aiming of the processing to assess aspects relating to a natural person[93] . As regards, on the other

[90] GPDP; Provision No. 488 of 24 November 2016. Web platform for reputational profiling, web doc. no. 579683.

[91] Court of Cassation, sec. I Civ. - 25/05/2021, no. 14381 .

[92] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.59

[93] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.60

hand, how the profiling process is implemented, doctrine is divided. According to some it is possible to configure profiling as a two-phase activity, while according to other doctrine it consists of three phases[94] . However, on a more in-depth reading that takes into account the constant change inherent in technological evolution, it is appropriate to consider that it is not necessary to consider the number of phases as static in every profiling process, and thus to consider a two- or three-phase process first and foremost, but it would be appropriate to consider that the phases of the process vary as the technological means employed change.
Already in Recommendation CM/Rec (2010)13 of the Council of Europe, point 1(*e),* there was a definition of profiling that already referred to statistical inference and the creation of categories or *clusters*. The Regulation, on the other hand, tends towards a greater closeness between the phase of profile creation and the predictive purpose for which it is directed[95] .
It should also be noted that, again referring to the Regulation, it contains the definition of profiling not in the technical sense but in the 'common and generally understood' sense[96] .
Moreover, the aforementioned legal act takes particular account of the protections for the data subject to be afforded by considering a *risk-based* approach. In fact, if profiling is carried out by means of automated processing - and also entirely automated, provided it does not produce significant effects - the ordinary rules of the GDPR will apply; if, on the other hand, processing is solely automated and produces significant effects, Article 22 of the Regulation[97] will apply.
To be read in conjunction with Article 4(4) of the GDPR are several other rules contained

[94] On this point see D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.61 and Pierucci, A. (2019). Data processing and profiling of individuals. Personal data in European law , p.424. In particular, D'Ippolito describes the biphasic doctrine in these terms: a first phase consisting in the collection of information about a person or a group of people; as well as the evaluation of characteristics and behavioural patterns in order to include them in reference groups or categories; a second phase consisting in the analysis of previously deduced information in order to predict a certain behaviour. In order for profiling to come into existence, according to this reading, it is necessary for both phases to take place; in fact, if processing stopped at the first phase alone, it would amount to a mere classificatory activity. With regard to the other reading, the second author considers that: "technically, profiling [can] be conceptually divided into three phases: 1. the phase of data warehousing, i.e. a phase in which observational data on the behaviour or characteristics of individuals are collected and stored on a large scale, and in which personal, coded, but also anonymous data may fall; 2. the data analysis phase that allows correlation between different behaviours/characteristics and other behaviours/characteristics (data mining); 3. an 'inference' phase in which, on the basis of certain behavioural variables or characteristics that are specific to an individual, new behavioural variables or characteristics are inferred."
[95] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.62
[96] On this point see more extensively D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.62. The author also discusses the concept of 'classification', which remains relevant under Art. 4(2) GDPR.
[97] See also D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp. 63-68. The author also refers to 'statutes of profiling', meaning different levels of protection with respect to the profiling phenomenon. In particular, four levels of protection are identified that look at the whole economy of the Regulation.

in the same act. First of all, Recital 24[98] of the same Regulation, which, after referring to the territorial scope of application, emphasises the 'monitoring of the data subject's behaviour', whereby it is appropriate to check whether natural persons are tracked on the Internet, including the possible subsequent use of personal data processing techniques (*first and foremost*, profiling techniques), in particular to take decisions concerning them or to analyse or predict their preferences, behaviour and personal attitudes[99] . Secondly, by reading Article 4(4) GDPR in conjunction with Recital 30[100] it is possible to understand even better the activity of online tracking, since the provision refers to identification tools of various kinds. Moreover, Recital 75 expressly refers to the danger of possible discrimination arising from the processing of data[101] . Last but not least, it is essential to consider the reading of Article 4(4) GDPR in conjunction with Article 22 GDPR for several relevant purposes. First of all, through said reading it is possible to identify various levels of data subject protection (see *above*); secondly, it is possible to analyse how traditional legal institutions - such as consent, identity, personality rights - are enriched, changed and challenged, while still preserving their traditional functions, by the advent of algorithmic technologies and Artificial Intelligence. In fact, as mentioned (see *above*), Article 22 GDPR deals specifically with automated processing followed by a decision solely based on the same, in order to regulate the autonomy of data subjects or profiled subjects. In particular, the rule gives the data subjects the right to intervene in the algorithmic process when the processing is entirely automated, i.e. does not involve any human input. In other words, the rule acts as a tool to minimise the risks that could be harboured by the results of mathematical and statistical processing carried out by predictive algorithms and artificial intelligence systems[102] . With respect to the structure of Article 22 GDPR, some doctrine has regarded it as a model rule[103] . The first paragraph of Article 22[104] recognises the data subject's right not to be subjected to decisions that are 'solely' based on automated processing, i.e. the data subject's right to have his or her claim to obtain a legal asset or, even more, his or her *status* not solely examined by a non-human entity. Indeed, the *discrimen* between partially automated and fully automated processing is given by the presence or absence of the human being within the decision-making process in the

[98] Art. 22 GDPR.
[99] Ciullo, M. Profiling and automated decisions: theoretical and evolutionary profiles, forthcoming.
[100] Recital 30) GDPR .
[101] Ciullo, M. Profiling and automated decisions: theoretical and evolutionary profiles, forthcoming.
[102] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.71. See also Art. 1 GDPR insert text
[103] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.72
[104] Insert first paragraph

substantive sense, i.e. of actual control[105] . Automated processing, as seen, also includes profiling as defined by the Regulation. In particular, when profiling takes the form of fully automated processing, it may fall within the scope of Article 22 GDPR[106] . A general prohibition against this form of processing follows from the rule, a prohibition that operates by default and which means that, by default and in the absence of action on the part of the data subject, wholly automated processing that produces particularly incisive effects on the data subject is prohibited[107] . On the notion of 'decisions based exclusively on automated processing [...] that significantly affect the [...] person (of the data subject)', as well as on particular categories of subjects and in increasingly advanced and sophisticated technological fields - such as the so-called 'Large Language Models' - the notion of 'decisions based exclusively on automated processing [...] that significantly affect the [...] person (of the data subject)' has been opened up. *Large Language Models* - a long debate has been opened, which was the starting point for the analysis and solutions put forward in this work. By 'legal effect' must be understood any processing that may impact on the exercise of a person's rights, freedoms or legal *status*; processing that has a similarly significant impact on a person must be understood instead as processing that does not merely produce changes in the rights and obligations of the person concerned, but rather produces relevant repercussions in the legal sphere of the same person that are worthy of protection. In this sense, profiling may be regarded as an example of processing producing relevant legal effects on the person .[108]

As regards, on the other hand, the second paragraph of Article 22 of the Regulation[109] , it lays down certain exceptions to the first. The doctrine[110] has classified them as contractual derogations, by provision of law and by explicit consent of the data subject. Fully automated

---

105 D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp. 72-73. See also Guidelines on automated decision-making concerning natural persons and profiling, cit.
individuals and profiling, cit., 23. The Author

106 D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.70. The author explains this assumption by clarifying that automated processing has a broader scope than profiling, and profiling techniques may not always constitute automated processing. Indeed, he continues, for example, 'inflicting a fine on the basis of the evidence provided by the speed camera is an automated decision-making process that does not necessarily involve profiling. However, if the decision to impose the fine stemmed from monitoring the person's driving habits over time, determining its disbursement and amount, one would have (entirely) automated processing based on profiling."

107 D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp. 74-75

108 D'Ippolito, G. (2021). Profiling and online targeted advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp.75-76

109 Insert standard

110 Pelino, E. (2023). The interaction between DSA and GDPR. In L. Bolognini, et al. (ed.), Digital services act and digital markets act: definitions and first applications of the new European regulations. Giuffrè, p. 268 reported by , D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.77

processing is permissible if it is based on lawfulness within the meaning of Article 6(1)(*b)* of the Regulation, i.e. a contract or the performance of pre-contractual measures: in this case, it must be demonstrated that the use of automated processing for the purpose of carrying out the activity is strictly necessary and not unjustifiably invasive of the data subject's legal sphere. Furthermore, fully automated processing is permitted in cases where it is provided for by a European Union regulation or that of a Member State, which also contains measures for the protection of data subjects. Lastly, the data subject may have recourse to the instrument of explicit consent - relying on the legal basis for the processing of special categories of data *under* Article 9(2)(*a)* of the Regulation - or in all other cases where the processing itself is shown to be particularly risky .[111]

The third paragraph of the provision in question[112] specifies that it will be necessary to adopt appropriate measures, which the data controller will identify in accordance with the principle of *accountability* set out in Article 5(2) GDPR, to protect the rights, freedoms and legitimate interests of the data subject in order to ensure that the purely automated processing is entirely lawful. In particular, the data subject is recognised the right to obtain human intervention by the data controller , the right to express his or her opinion and the right to contest the decision produced by the algorithm[113] . *The leitmotif* that characterises the various protection measures remains, however, that of giving human input a decisive role, which in some way curbs the concern that had moved the 1995 legislator, while granting great importance to the benefits of automation.

Finally, the fourth paragraph of the aforementioned provision contains a tightening of the discipline, the *rationale for which* lies in the greater and potential risks that profiling and automated processing could bring about, since they are based on sensitive data and inherent to the most intimate aspects of the data subject's life[114] . In particular, the rights that constitute the pillars of the algorithmic and 'de-humanised' *due process* are included in Article

[111] D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, p.78. The author also argues about the limits of the rule in question: it would, in fact, be a rule that does not incorporate particularly stringent guarantees for the data controller, nor be an absolute novelty of the Regulation. Moreover, a major limitation of the provision at issue here would seem to be that it would greatly erode the scope of paragraph 1. Finally, there would be a legal vacuum in some respects, since automated processing sic et simpliciter would be subject to the same discipline as partially automated processing.

[112] Insert rule

[113] On this point, see more at length D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp. 80-83.

[114] More on this point, see D'Ippolito, G. (2021). Profiling and targeted online advertising: real-time bidding and behavioural advertising. Edizioni scientifiche italiane, pp.83-84

22 GDPR[115] . The provision stipulates that 'the data controller shall implement appropriate measures to protect the rights, freedoms and legitimate interests of the data subject, at least the right to obtain human intervention on the part of the data controller, to express his opinion and to contest the decision'.

Although the legal provision does not explicitly refer to the right to know the reasons behind the automated decision, its existence, in terms of positive law, appears unquestionable. In fact, a systematic and axiological reading of Article 22 allows this prerogative to be included among the fundamental rights, while guaranteeing its effectiveness[116] . This brings about a significant point of convergence between ethical and legal discourse on artificial intelligence. This conclusion is based on a double interpretative binary, which - although it can be integrated - retains, from an argumentative point of view, full autonomy: on the one hand, that of the fundamental rights of the human person; on the other, that of the functional logic within Article 22 GDPR, considered as the regulating norm of the algorithmic process relating to the data subject.

From the point of view of legal values, the interest of the data subject in knowing the reasons for the decision clearly emerges: it stands in close correlation with pivotal principles of the European democratic constitutional order, expressly referred to at the heart of the GDPR's structure, i.e. transparency and the protection of personal dignity. In the overall architecture of the Regulation, these principles are closely interconnected: transparency is, in fact, an essential element of personal data protection, i.e. of that system that protects data in order to protect the person to whom they refer .[117]

3.3. Profiling in subsequent European legislation.

Following the introduction of the General Data Protection Regulation, which represented a fundamental point of reference for the regulation of personal data processing and profiling in the European Union, other European regulations and legislative acts have continued to address the issue of automated profiling, often in specific contexts. First of all, the Digital Services Regulation (DSA in its English acronym), approved as EU Regulation 2022/2056,

[115] Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021), p.VIII.
[116] Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021) p.VIII.
[117] Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021), p. VIII.

should be mentioned. The analysis of the points of intersection between the two regulations under consideration - the General Data Protection Regulation and the Data Services Act - is crucial not only in itself, but also to understand how the *enforcement* mechanisms of the two regulatory instruments may interact or conflict, making both substantive and procedural coordination between them necessary. With respect to the issue of profiling, several provisions contained within the DSA come to the fore. First of all, the art.26, par. 3 DSA[118] which places a ban on advertising to recipients based on profiling - as defined by the GDPR - using the 'special categories of data' referred to in art.9, par. 1 GDPR[119] . With reference to 'special categories of data', the case of *Vyriausioji tarnybinės etikos komisija*[120] should be considered. Again, Article 28 DSA[121] introduces a further prohibition on the presentation of advertising based on profiling if the online platform provider knows, with reasonable certainty, that the recipient of the service is a minor. With regard to this category of users, targeted advertising is absolutely prohibited. Furthermore, Article 38 of the DSA[122] imposes an additional restriction for large online platforms (VLOPs): if they use recommendation systems, they must guarantee recipients at least one option that is not based on profiling. In general, the demands that moved the legislator of the *Data Services Act* partly resemble those of the previous legislator of the *General Data Protection Regulation*. Once again, we were faced with the redefinition of the digital ecosystem, in this case also with specific reference to the *media*, with respect to the legal system and framework. The attempt to unravel the apparent neutrality of the information intermediation operated on platforms has been crucial, as well as - and this is particularly relevant with reference to profiling activities - confronting the consequences of the pervasiveness with which platforms can control, shape and customise information algorithmically, through predictive analysis based on behavioural data[123] . The Digital Services Act (DSA) represents a turning point in the regulation of the European digital ecosystem, introducing for the first time an *ex ante* prescriptive approach to the management of online information circulation. In this new regulatory framework, the *safe harbour* principle - already provided for in the Electronic Commerce Directive (2000/31/EC) - is reworked, attributing a form of legal liability to platforms in relation to illegal or harmful content conveyed through their services. The declared objective is to create a safer transnational digital space in which users' fundamental rights

[118] Art.26, para. 3 DSA.
[119] Pelino, E. (2023). The interaction between DSA and GDPR. In L. Bolognini, et al. (ed.), Digital services act and Digital markets act: definitions and first applications of the new European regulations. Giuffrè , p.12
[120] CGUE, C-184/20, 01/08/2022 .
[121] Art. 28 DSA
[122] Art. 38 DSA
[123]DE VIVO, I. (2024). The power of opinion of online platforms: what role of the European 'regulatory turn' in the digital information oligopoly? *Federalismi. en*, *2*, 45-75, p.263.

are guaranteed and protected.

However, a critical examination of the provisions highlights the difficulties encountered by the DSA in attempting to limit the enormous power wielded by large digital platforms, especially those of an infrastructural nature. These obstacles derive, to a large extent, from the application of a traditional, vertical (*top-down*) regulatory model to a highly dynamic, multilevel digital environment, such as that of today's *platformed internet* .[124]

Article 16 of the DSA introduces - for the first time on a European scale - a structured *notice and takedown* mechanism, which has already been tested in national contexts such as the NetzDG law in Germany (2017) and the Loi Avia in France (2020). This provision configures a kind of active surveillance obligation: platforms must swiftly examine reports of allegedly illegal content and, if necessary, proceed to its removal. However, the task of defining what is to be considered illegal is left to the laws of the individual Member States, resulting in a regulatory fragmentation that risks undermining the goal of harmonisation at European level. Moreover, it is not immediately clear whether these provisions apply only to legally illegal content or also to content that is legitimate but deemed inappropriate according to the internal rules of the platforms. A combined reading with Article 17 suggests that takedown may also be justified in the case of lawful content that violates contractual terms of use, effectively widening the scope of intervention of platforms and including both categories of content[125] . What has been illustrated falls within the ranks of the issue of profiling, since explicit moderation, analysed through the reading of the provisions of Article 16 DSA, is not the only form of moderation that characterises neo-intermediation. The other form, less explicit but equally incisive, is that exercised through algorithmic systems of personalisation and profiling, in which a central role is played by the so-called Recommendation Systems (RS). These software tools, based on the analysis of personal data and user metadata, are designed to suggest targeted and individualised content. In many cases, RSs operate by creating associations between users with similar behaviour or characteristics, thus generating homogeneous groups built according to criteria known only to platform managers. By means of advanced profiling techniques, such as *psychography* or so-called *hypernudging*, such systems can not only get to know users in depth, but also influence their choices through targeted, potentially manipulative communications. The result is a fragmentation of the public that weakens the critical function traditionally exercised by the autonomous public sphere. According to the Habermasian conception, in fact, a healthy

[124] Flew, T. (2019). The platformized internet: Issues for media law and policy. *Available at SSRN 3395901*.
[125] DE VIVO, I. (2024). The power of opinion of online platforms: what role of the European "regulatory turn" in the digital information oligopoly?. *Federalismi. it*, *2*, 45-75, pp. 264-265.

public sphere should generate legitimising power through confrontation and communication, thus maintaining a democratic balance vis-à-vis central institutions. In this context, making the functioning of RS transparent is an essential step to reduce the information asymmetry between users and platforms. The European legislator seems to have recognised this need. Although it has not fully accepted the recommendations of the European Data Protection Supervisor (EDPS), which called for a general ban on profiling in commercial and political advertisements, Article 27 of the Digital Services Act introduces a number of provisions aimed at improving the transparency of recommendation systems. In particular, it calls on the *Very Large Online Platforms* (VLOPs) to clarify the criteria used for profiling and to offer users the possibility of choosing between different modes of personalisation.

However, although it represents an important step towards greater transparency, the standard is not fully satisfactory. Indeed, there is no concrete indication as to which options should actually be made available to users and how these can be compatible with the protection of fundamental rights and democratic values. Moreover, there is no obligation for platforms to allow the adoption of alternative and independent, potentially more neutral recommender systems.

In the absence of strict constraints, it is plausible that RSs continue to reflect - and reinforce - the cognitive biases, conscious or otherwise, inherent in the logics of the platforms in which they are embedded. The implications of this for the protection of fundamental rights and the autonomy of the public sphere are already widely recognised. The pervasiveness of personal data collection and analysis, fuelled by surveillance-based business models, seems destined to continue unabated.

In doing so, the legislation risks neglecting an essential element for a democratic governance of the digital environment: the re-establishment of what we might call *intellectual privacy*, i.e. the possibility for each individual to exercise effective control over the flow of information concerning him or her, both incoming and outgoing. Only by guaranteeing this form of cognitive self-determination will it be possible to counter the dominant narratives and normalisation mechanisms imposed by both large technology companies and institutional apparatuses .[126]

Another act worth mentioning is EU Regulation 2022/868, otherwise known as the Data Governance Act or DGA. It lays down rules on *data sharing*, and in this case the reference to

[126] DE VIVO, I. (2024). The power of opinion of online platforms: what role of the European "regulatory turn" in the digital information oligopoly?. *Federalismi. it*, *2*, 45-75, p.267-268.

profiling takes on a peculiar role. In fact, the Regulation never expressly refers to profiling, but it seems clear that the ethical and responsible use of data advocated in the act may refer across the board to phenomena related to profiling .[127]

4. Classifications, debated issues, problems underlying profiling and automated processing.

With regard to the scope of profiling and automated decisions, several problematic issues have arisen. First of all, there is an issue concerning the link between the human person and his or her data, i.e. the dichotomy between personal identity and digital identity. The link between the person and his or her data is significantly weakened due to the complexity of the new processes of personal identity construction. These processes are configured - to a certain extent - as 'separate' from the individual, to the extent that they are entrusted to the computational power of the technological apparatus and, consequently, to the self-referential dimension of the computer-statistical code that governs it. It is the machine, in fact, that recomposes the identity of the individual, which is fragmented in the multiplicity of data that represent him, and generates - from these data - new personal information through the techniques of *big data analytics*. The repercussions at the level of fundamental rights are obvious: the intangible core of the protection of the human person necessarily includes - and it could not be otherwise - control over the work of the machine, as an expression of guarantee of respect for the dignity and autonomy of the individual in the digital context[128] . From a legal perspective, exercising control over the image of the person processed by an automated system is equivalent, in substantive terms, to exercising control over the motivations behind the decisions generated by the machine. This equivalence is evident in the light of Article 22 of the GDPR, in which personal identity takes on a dual role: on the one hand, it represents the final recipient of the decision - which, by definition, must produce significant legal effects or in any case significantly affect the individual's personal sphere -; on the other, it constitutes the basis on which the decision is justified, by means of an assessment of personal elements concerning the individual concerned, such as his or her aptitudes, economic conditions, state of health, preferences, behaviour and other relevant aspects. This dual function clearly shows that control over the automated decision necessarily implies control over the process of reconstruction - hetero-directed and technologically mediated - of personal

[127] See on this point, Bravo, F. (2022). 'Recipient' of information and processing of personal data in the evolution of the European legal system. In M. D'Auria (ed.), I problemi dell'informazione nel diritto civile, oggi. Studies in honour of Vincenzo Cuffaro (pp. 431-454). Roma Tre-Press and Bravo, F. (2021). Personal data brokering and data sharing services from GDPR to the Data Governance Act. Contract and Enterprise Europe, 1(1), 199-256.

[128] Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021), p. IX.

identity. In this context, it is essential to understand the logic behind algorithmic outputs: this is the only way to guarantee effective protection of the individual. From a legal point of view, knowing the reasons for the decision does not simply mean knowing its outcome, but understanding the mechanisms through which the identity of the person concerned was reconstructed and how this reconstruction affected the outcome itself. The example of the personal profile clearly demonstrates this: it is not enough to have access to it, it is necessary to understand its genesis, the reasons for its attribution and its impact on the final outcome. At the level of fundamental principles, the right to know the reasons for the automated decision is based on the informational and computational nature of personal identity in the contemporary digital context. It represents the essential content of the individual's power to supervise the processes of technological hetero-construction of one's identity. From this perspective, control over the machine is configured as a new dimension of privacy, understood as freedom from the imposition of arbitrary and non-transparent identity models, as well as a central element of a democracy capable of exercising effective control over digital power. In internal key to the GDPR, the discourse finds an even more direct formulation in Article 22, which regulates automated decision-making processes in terms of a 'de-humanised' construction of the digital person, while providing for a series of minimum guarantees (para. 3), including: the right to human intervention, to express one's point of view, and to challenge the decision. These guarantees delineate an individual's power to scrutinise the algorithmic process and reveal the essential purpose of the regulatory intervention: to ensure compatibility between automated data processing and the protection of fundamental rights, as enshrined in Article 1(2) of the GDPR. The connection between data and personal identity is evident here: if an individual's identity is made up of his or her data, any intervention on the data translates into an intervention on the person himself or herself.

It follows that effective control must necessarily include the possibility of understanding the 'reasoning' behind the output generated by the machine. If this logic remains unknown, two consequences appear inevitable: human intervention, envisaged as a remedy, is reduced to a mere formality; and the final decision cannot be substantially challenged, lacking access to its reasoning.

The structure of Article 22, read in its functional dimension, thus suggests an interpretation oriented towards transparency as 'readability', i.e. as the possibility for the person concerned to understand the internal logic of the decision-making process that concerns his or her identity. This is, of course, the transparency needed to guarantee effective control over the

decision. Hence the relational dimension established by the Regulation between data controller and data processor: the former has the right to receive from the latter a meaningful explanation of the decision based on the assessment of his identity. The limit to this obligation of transparency does not coincide with a total disclosure of the algorithm, but requires the disclosure of that information that is essential for the protection of the person involved in an algorithmic process of construction of his identity.

Ultimately, from the traditional point of view of the protection of personal rights, the centrality of a subjective power of control over the use of one's own data within automated decision-making processes clearly emerges. This power reflects and protects fundamental constitutional values, such as individual self-determination and human dignity. From a more current perspective, attentive to the transformations of the infosphere and to the risks of a new form of domination exercised through computational power, the right to understand the workings of the machine acquires further significance: it represents a tool for opening up the algorithmic 'black box' and a defence of human freedom in the digital context.

However, there is no shortage of those who contest the effectiveness of this right. According to some critics, although legally enshrined, the right to an explanation would be difficult to realise in practice. From a technical point of view, it is observed that *autonomic computing* models would, by their very nature, be inaccessible in the crucial phase in which they determine outputs. This would not only be a question of communication between different cognitive systems - the human and the artificial - but rather a structural limitation: in machine learning systems, the operating algorithm often remains unknown even to its developers, making it impractical to understand its functioning in the decision-making phase. Indeed, the explanations currently developed in computer science are not designed to meet the needs of the GDPR, but aim to ensure the efficiency of predictive computing, rather than the regulatory transparency required for the protection of rights. In other words, ensuring the intelligibility of machine decisions would inevitably entail a sacrifice in terms of technological opacity. This tension illuminates a fundamental issue: if the predictive value of the algorithm is deemed sufficient to compensate for the inaccessibility of its inner workings, a dangerous asymmetry looms between the machine's ability to understand the human being and the latter's inability to understand the machine. The result would be a computational identity totally opaque to the same subject to which it belongs, with the consequent loss of any possibility of exercising control over the process of identity formation. In the final analysis, the power to know the logic and reasons for the automated decision, as envisaged by the GDPR, risks proving inadequate with respect to the reality it intends to regulate. The

paradigm of transparency could, in this context, prove to be inappropriate, generating - paradoxically - new opacity in the digital environment it purports to regulate .[129]

Changing perspective, but in any case considering the issues that find their genesis in profiling and automated decisions, the older doctrine, from a technical point of view, has discussed whether, notwithstanding the letter of the provision of Article 22 GDPR that speaks of "the right not to be subjected to ...", the Regulation actually intended to attribute a "subjective *right*" (*right)* to the data controller or, instead, as is mostly believed, to establish a *prohibition* (*prohibition*) to which the data controller is bound, regardless of any specific initiative of the data subject. In particular, this school of thought has considered the configurability of a general prohibition to be more in line with both the general *ratio of* the Regulation to increase the protection of data subjects, and with the principle of *accountability*, which requires the data controller to take an active and diligent part in organising the data processing processes so that they are demonstrably respectful of the rights and freedoms of data subjects. In fact, making the effectiveness of the restrictive principle depend on the latter taking the initiative to exercise their supposed right, would risk, given the speed, complexity and scale of the processes for which this discipline is deemed to be most significant, depowering that very principle to the point of rendering it effectively inoperative, except in exceptional cases. Again, with regard to the expression '*a decision [...] that affects his person in a similarly significant manner*', problems arise not only of a technical nature but also and above all of an ethical nature. In fact, the vagueness and generality of the expression could act as an open clause through which all concerns about the potential discriminatory nature of algorithms could be addressed, paradoxically offering the guarantees and protections provided by the rule in question. It is true that many decisions are exclusionary, in the sense that they may significantly affect the recipient's life, depriving him or her of certain opportunities (the examples are various, one can refer to access to education, health services, credit); however, for these very reasons, the GDPR requires a significant human presence. Thus, 'human intervention should in these cases also be concerned with assessing issues related to the discriminatory or unfair potential of algorithmic processes as well as protecting the dignity of the recipients'.

Another profile of particular interest is that relating to so-called *nudging*. This term refers to "those modes of conditioning towards certain choices that do not, however, exclude the possibility of different options, even if these are somehow discouraged, for example, by the

[129] Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021), pp. IX-XIII.

way they are presented[19] ". In this regard, it can be argued that such techniques are not included in the prohibition of Article 22 GDPR, however, a more restrictive reading should admit the possibility of including those cases where they give rise to significant consequences 'similar' to legal effects for the subjects. With regard to profiling, on the other hand, one of the most insidious issues could be the creation of clusters, i.e. classes or groups of individuals created by the algorithm on the basis of patterns or aspects in common among the subjects analysed. Clearly, being placed in a group because of certain habits is a relevant phenomenon when the categorisation of the individual in one group rather than another involves a decision that affects the enjoyment of one's rights. Of course, as some of the doctrine emphasises, such automated decision-making processes pose several challenges for the freedoms of individuals, especially for the risks arising from discrimination due to the categorisation of individuals into more or less homogeneous groups. Profiling and the automated decisions based on it constitute processes whose discriminatory scope is implicit since they tend to group groups of individuals into categories on the basis of common characteristics and make decisions on the basis of belonging to a specific group[130] . Many of the issues have been resolved in subsequent legislative acts. This is, for instance, the case of so-called *nudging*.

5. Profiling and technological evolution: the European Union Regulation on Artificial Intelligence.

Faced with a digital and technological reality that is constantly changing and evolving, the European Union has for some time now laid the foundations and approved an *ad hoc* regulation on Artificial Intelligence. This was a breakthrough of global dimensions, Regulation (EU) 2024/1689 being the first piece of legislation in the world to dictate rules concerning Artificial Intelligence. The landmark piece of legislation underwent a fairly lengthy process that led to its final approval only on 21 May 2024. It was preceded in 2020 by a White Paper formulated by the European Commission entitled 'White Paper on Artificial Intelligence - A European approach to excellence and trust'. This marked the beginning of intense discussions and debates within the European institutions in October 2020, leading to a first official proposal for an AI Act on 21 April 2021. On 6 December 2022, the Council of the European Union then adopted the general approach, which allowed negotiations with the European Parliament to begin. On 9 December 2023, after three days of heated debates and already entering the historic and salient moments of the supranational

[130] Ciullo, M. Profiling and automated decisions: theoretical and evolutionary profiles, forthcoming.

order, what appeared to be the final version of the AI Act was reached. A subsequent final version of the text was unanimously approved by the EU Member States on 2 February 2024. On 13 March 2024, the AI Act was finally passed by the EU Parliament, marking the advent of the first global regulation on artificial intelligence. The Regulation entered into force on 1 August 2024[131] . With respect to the phenomenon of profiling, it refers to Art.4 para.4 GDPR as stated in Art.3 No.52 AI ACT. Profiling then appears in several Recitals, which open the legislation. *First of all*, Recital 10[132] refers to profiling by dealing with the fundamental right of individuals to the protection of personal data, which remains firm also in the regulation we are dealing with here; Recital 42)[133] states that natural persons should never be 'judged on the basis of AI conduct based solely on profiling'; Recital 53)[134] focuses on the risks associated with AI systems that involve profiling as defined by European laws, such as the GDPR, and related directives and regulations; Recital 59)[135] instead emphasises profiling when it is used to determine the risk of crime or recidivism, taking into account the personality traits and past criminal behaviour of individuals or groups. Finally, Article 5(1)*(d)* of Regulation[136] mentions profiling by referring to prohibited AI practices: in this sense, the provision states that it is prohibited to use artificial intelligence systems to make assessments of individuals on profiling or on the evaluation of personality traits and characteristics.

6. From profiling to personalisation.

Reading the combined provisions of the rules enunciated so far and comparing the case law of national and supranational matrix on the subject of profiling, it seems to emerge, at a closer look, an adjacent but different phenomenon from mere user profiling. This is the so-called personalisation: on the one hand, it appears to be the natural consequence of the profiling activity and, to some extent, also the result of certain automated decision-making activities; however, on the other hand, it is clear to what extent this phenomenon can constitute an autonomous category, aimed at affecting the protection of the rights of the individual. As already illustrated, the legislation defines profiling by indicating, in essence, different but necessary stages for profiling to take place. In particular, reference is made to

[131] Ciullo, M. (2024, October). Large Language Models: Ethics and Norms in the European Union. In 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE) (pp. 1065-1070). IEEE.
[132] Recital 10) AI ACT
[133] Recital 42) AI ACT.
[134] Recital 53) AI ACT
[135] Recital 59) AI ACT
[136] Art. 5 para. 1(d) EU Regulation 2024/1689

the use of personal data collected for the purpose of assessing personal aspects of the natural person and, in this case, to analysing and predicting the behaviour of that person. Article 22 of the GDPR, for its part, emphasises that the data subject has the right not to be subjected to an automated decision concerning him or her, including profiling. Similarly, as analysed, the subsequent legislation also dwells on different declinations of profiling and automated decisions, sometimes resolving problems and *vexatae quaestiones* that had arisen with reference to the two phenomena, but the legislation neither defines nor regulates the phenomenon of personalisation. It is a phenomenon that takes its cue from profiling and automated decisions, being a consequence of the first two, however, it differs from them. In attempting to define the phenomenon, it does not seem wrong to first consider the linguistic definition of the term[137] . From a linguistic point of view, all definitions converge on the imprint or centrality not so much of the individual, but of the characteristics of the individual. From a legal point of view, although there is as yet no normative definition of the term, the phenomenon of personalisation can be understood as the result of automated decision-making activity, including and above all profiling activity, which assumes legal relevance for various reasons. Moreover, in the concept of 'personalisation', considered in its legal meaning, it is easy to bring back many of the controversial issues already discussed (see *above*). In this sense, it is easy to see how so-called. Surveillance capitalism leaves room for different categories, which nevertheless find their genesis in it. It would seem, in fact, more appropriate to refer to a capitalist system that not only monitors the user/person but which, as a result of the acquisition of data related to them, manipulates their choices or replaces their faculty of choice *tout court*. It is clear that in this context, the self-determination of the individual is increasingly reduced, to the point - probably - of disappearing altogether in certain circumstances. In general, it can be said that the scope of the concept and meaning of self-determination, a fortiori considered according to traditional legal doctrine and in any sphere of private law that is confronted with the digital and automated set-up, faces a profound crisis precisely because platforms and Artificial Intelligence cause users to adopt choices that are not entirely conscious, sometimes even dangerously undermining the fundamental

[137] The Treccani dictionary under the headword "Customisation" reads: "The act, or the fact, of customising, of being customised (in the various meanings of the verb); in particular, attributing a character, a personal imprint to objects that would otherwise be banal, commercial, mass-produced". The Garzanti dictionary: "1. to adapt to the tastes, to the needs of one or more people, of a clientele: to personalise one's car; to customise an environment, a product, software
2. consider as personal: personalise political contrasts'.

principle of human dignity[138] . It is precisely the characteristics of a *platform economy* that is increasingly moving in a distinctly (before the advent of the AI Act, and now more) veiledly manipulative direction that have been the spur for the European legislator to regulate in order to control the growing power of large private companies[139] . The use of artificial intelligence to monitor and track the preferences and surfing behaviour of users within closed digital environments raises relevant questions in terms of the protection of fundamental rights. It is undeniable that the expansion of digital platforms has had a positive impact on the circulation of information, widening opportunities for access to knowledge and making content of all kinds available very quickly. In this context, the figure of the so-called *digital prosumer* asserts itself: a user who not only consumes information, purchases goods and uses online services, but also actively participates in the creation and sharing of content. This apparent protagonism of the user in the digital landscape has fuelled the idea that greater availability of data and information sources automatically translates into greater awareness and rationality in the choices made online. However, this belief turns out to be largely illusory when the facts are tested. The algorithmic customisation mechanisms and disintermediation processes operated by platforms do not increase user freedom, but rather simulate it. In reality, it is precisely the architecture of digital systems - designed to offer content selected on the basis of invisible and highly personalised criteria - that generates the impression of autonomous control and choice free of conditioning. The user thus finds himself interacting in an environment that, while seemingly open, is profoundly shaped by opaque logics that influence what he sees, reads and spends his time on. Digital platforms have introduced a new imbalance in the

[138]Sposini, L. (2024). Democracy in the Age of Digital Platforms between Profiling and Content Personalisation. The Paradox of Unconsciousness, pp. 279-280. See also, more extensively Quarta, A., & Smorto, G. (2020). *Private Law of Digital Markets.* Le Monnier University, p. 147.

[139] See, more extensively, Sposini, L. (2024). Democracy in the Age of Digital Platforms between Profiling and Personalisation of Content. The Paradox of Unconsciousness. p. 280. The author warns that "Most of today's digital markets exhibit a naturally anti-competitive and monopoly-prone structure, and this is basically due to the presence of 'network externalities' and 'economies of scale'. The first expression refers to the fact that in the digital market, the usefulness of a certain intermediation service depends on the number of other actors using it, as is the case, for instance, on social networks: here, in fact, the user will only be prompted to register where there is a large number of members. However, for some platforms (in particular marketplaces such as Amazon), what is realised are rather 'cross-network externalities', whereby the user's interest does not increase according to the number of other members, but by virtue of the growth of those operating on the other side of the platform: as the number of sellers using Amazon's services increases, so will the number of potential buyers. Network externalities then favour concentration because they make users dependent on the platform by building an entire closed integrated system where they have all the services and products they need at their disposal. This strategy increases exit costs to another platform and thus produces a lock-in effect that in turn strengthens the platform's own dominant position in the market. A further element that favours this 'platform capitalism' is the presence of returns of scale due, in particular, to their ability to collect and process an infinite amount of data, both those consciously provided by users and those derived from tracking their online activity. These relationships are clearly unbalanced in favour of digital agents'.

relationship between economic power and citizen, radically redefining traditional roles. The user is no longer just a consumer of the products and services offered by large technology companies, but also assumes a dual function: on the one hand, he becomes a co-producer of value through content, reviews and interactions; on the other, he himself becomes a commodity, as the data generated by his online activities are systematically collected, processed and often monetised.

This complex dynamic takes place in a largely invisible manner, so much so that the individual tends to consider himself fully autonomous in his choices, ignoring the profound - and often opaque - role of the digital infrastructures that filter and determine the information content he has access to. In other words, the perception of freedom in surfing, getting information or communicating is strongly influenced by mechanisms that select and propose contents on the basis of non-transparent algorithmic criteria. In other words, if on the one hand the Internet, through its platforms, amplifies the consumer's freedom by providing him with what he wants, on the other hand it ends up compressing the citizen's freedom, depriving him of the information he needs to form a critical opinion. The result is a closed information reality, in which everyone remains confined within their own worldview, constantly confirmed and never questioned, with a substantial impoverishment of the capacity for autonomous and pluralist thought[140] . So much so that, also based on these reflections, as well as on the development of *neuromarketing*, the particular branch of so-called *Artificial Emotional Intelligence* was born, which aims to programme artificial intelligences so that they have the capacity to interpret human emotional reactions, to calibrate their *output* with respect to these, and to influence the user so that he or she feels certain emotions[141] . On this last point, in particular, European Union Regulation No. 2024/1689 has intervened, which, in legislating on Artificial Intelligence, has banned manipulation, among other things; however, the issue remains controversial (see *below*). The reflections carried out so far clearly show how digital platforms and Artificial Intelligences, through the mechanisms of content profiling and personalisation, can constitute a concrete threat to some of the founding principles of democratic societies, first and foremost individual autonomy. Precisely because content is no longer the result of a free and open choice, but is selected on the basis of behavioural traits and inferred preferences, of data collected through different modalities, the user finds himself exposed to a customised representation of the world, which ends up unconsciously orienting his opinions, desires and choices. This dynamic, apparently neutral or functional to the '*user experience*', actually has a profound

[140] Sposini, L. (2024). Democracy in the Age of Digital Platforms between Profiling and Content Personalisation. The Paradox of Unconsciousness, pp. 280-281; the author, on this point, recalls Sunstein, C. R. (2018). Republic: Divided democracy in the age of social media.

[141] Sposini, L. (2024). Democracy in the Age of Digital Platforms between Profiling and Content Personalisation. The Paradox of Unconsciousness, p. 282. See also McStay, A. (2018). Emotional AI: The rise of empathic media.

impact on the individual's freedom of thought and self-determination. If the individual loses the ability to form an autonomous opinion and to critically confront different perspectives, the very possibility of participating authentically in the public debate is compromised, as is the possibility of making informed choices in the relevant legal system. less relevant are the implications for another pivotal value: human dignity. Understood not only as respect for the person, but as the foundation of autonomy and the ability to make conscious decisions, dignity risks being debased in a digital and automated environment that treats the individual as an object of predictive analysis, rather than as an active subject of his or her own informational destiny.

Chapter II

'Problematic Profiles of Profiling. The question of vulnerability'.

SUMMARY: 1. The key to understanding the *vulnerable individual*. The vulnerability of individuals in the context of personal data protection. 2. Digital vulnerability in European private law. 3.Limitations and alternatives to the *vulnerability-based* approach.

The key to understanding the *vulnerable individual*. The vulnerability of individuals in the context of personal data protection.

Recent advanced machine learning capabilities, coupled with automated decision-making processes, hyper-personalisation of behavioural advertising and predictive analytics, significantly contribute to widening the power gap between individuals and the entities that manage and process data in the digital ecosystem. This growing power imbalance has profound and multidimensional implications. In particular, the effects of such an unfair and disproportionate imbalance can be observed from different analytical and practical perspectives. One critical aspect that clearly emerges is the issue of discrimination, which takes on new and complex connotations in this advanced technological context[142] . This is why it is of great interest to think about the key of interpretation offered by a *vulnerability-based* approach[143] . Indeed, the entire human condition is becoming increasingly transparent and almost fragile in the face of a compact, opaque, pervasive and elusive power .[144] *The vulnerability-based* approach lends itself well to the analysis and application of norms relating to phenomena that are still partially analysed and regulated, which concern, on the one hand, profiling and automated decisions (see *above*) and, on the other, the challenges posed by Artificial Intelligence (see *below*). In particular, according to this doctrine, it would be desirable to reconceptualise the notion of vulnerability in the context of personal data protection, thus proposing a contextual, relational and layered definition that can effectively respond to the challenges posed by the algorithmic environment and ensure adequate protection of individuals' fundamental rights. This objective stems from the perceived lack of clarity regarding the vulnerability of data subjects within the General Data Protection Regulation (GDPR). In particular, it is emphasised that although the GDPR explicitly identifies only one

[142]Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. Calif. L. Rev., 104, 671.; Wachter, S. (2020). Affinity profiling and discrimination by association in online behavioral advertising. Berkeley Technology Law Journal, 35(2), 367-430.

[143] Malgieri, G. (2023). Vulnerability and Data Protection Law. Oxford University Press. This volume was used as a model for the entire structure of chapter II and the contents therein refer to it.

[144] See Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021).

vulnerable category, namely children, the entire data protection framework is implicitly based on a broader, layered and contextual conception of vulnerability. This implicit conception requires explication and deepening to adequately address the complexities introduced by the increasing use of algorithmic technologies[145] . Similarly, this approach could be adopted to critically read the recent EU Regulation on Artificial Intelligence, which, although it contains explicit provisions on the dangers of discrimination, also presents problematic profiles with reference to so-called vulnerable individuals or vulnerable categories. Generally speaking, the issue is based on a crucial knot of change resulting from the informatics-statistical government of reality: the human being's condition of vulnerability is the inevitable cost in the face of technological efficiency. A condition that is opposed to the constitutional and democratic axiological context of reference[146] . First of all, in this regard, it is necessary to identify a normative definition of an individual[147] , before that of a vulnerable individual. Under the aegis of the GDPR, it is possible to identify four requirements that define, albeit indirectly, the data subject. In particular, the latter is defined when defining the meaning of 'personal data' in Article 4(1)[148] . More explicitly, the data subject is the identified or identifiable natural person to whom the personal data relate. At least three details can be deduced from this definition: it is a *natural person* and not a legal person; some data must be *linked to* it; it must be *identified or identifiable*[149] . More generally, it can be said that the notion of Data Subject is in principle general, inclusive, unique and universal[150] . The first requirement inferable from the scope of the GDPR refers to the *living person*. This leads to

[145] Article 29 Data Protection Working Party on the Protection of Individuals with regard to the Processing of Personal Data (WP29), Guidelines on Transparency, WP260rev.01, 10; WP29, Opinion on legitimate interest, WP217, 40- 41; WP29, Opinion on purpose limitation, WP203, 25 and 32; WP29 Guidelines on DPIA, WP248, 9; WP29, Guidelines on Profiling, WP251rev.01, 22. See also EDPB, Opinion 3/ 2019, 23 January 2019, 6.

[146] Messinetti, R. (2021). Communicating in the infosphere. The vulnerability of the digital person. *FEDERALISMS. IT*, (18 of 28/07/2021), p. XVI.

[147] With regard to the definition of the individual, consider more extensively Malgieri, G. (2023). Vulnerability and Data Protection Law. Oxford University Press, Chapter 2.

[148] Art. 4 par. 1 GDPR

[149]Malgieri, G. (2023). Vulnerability and Data Protection Law. Oxford University Press, p. 18. On this point, also keep in mind WP29, Opinion 4/ 2007 on the concept of personal data, Adopted on 20 June 2007, WP136, 21.: with reference to the definition of natural person the Working Party 29 specified that, for the purposes of the application of the Data Protection Directive, a natural person is a living human being. The subjectivity of data cannot be limited to citizens or residents of a particular country; in other words, the right to the protection of personal data is universal and is not limited to citizens or residents of a particular country. Indications concerning the definition of the data subject and, more generally, the natural person referred to in the GDPR, are also provided by Recital 4 and Recital 2 of the GDPR: the former emphasises that automated data processing must be designed for the benefit of man, the latter emphasises the protection of the freedoms and fundamental rights of individuals. Furthermore, reference can be made to Article 6 of the Universal Declaration of Human Rights, according to which everyone has the right to be recognised everywhere as a person before the law.

[150] The Council of Europe had already expressed itself in the same terms with the Convention 108 for the Protection of Personal Data of 1981 : one agrees with this universalistic conception of the data subject, that everyone is equally potentially vulnerable.

the inevitable question of conceived but not yet living persons and that of deceased persons. On this point, the WP29 warned to take into account the transposition by the individual Member States, which are given the possibility to regulate data protection rights for unborn or deceased subjects[151] . The other aspect to be considered when speaking of a 'living natural person' is the *status of* the unborn child. The WP29 states that national civil law may have different rules on the legal *status* of the unborn, but that this is outside the scope of EU data protection law. Consequently, the general approach of the various national legal systems must be considered, 'together with the idea that the purpose of data protection rules is to protect the individual'. In any case, it can generally be said that their subjectivity in data protection law depends on national rules and specific situations and is therefore contextual. The second requirement of the definition of data subject in the GDPR refers only to natural persons and not also to legal persons, from which it follows that the latter cannot have the status of data subject. In fact, the European data protection framework has not always excluded legal persons from the scope of data subjects. Convention 108 of the Council of Europe of 1981 protected 'any individual', without mentioning legal or natural persons. Furthermore, Article 3(2)(*b)* of the Convention specified that any state could notify 'that it will also apply this Convention to information concerning groups of persons, associations, foundations, companies, firms and any other bodies constituted directly or indirectly by individuals, whether or not such bodies have legal personality '[152] . While the GDPR ensures, in Recital

[151] Several commentators have argued that although in principle the concept of privacy should apply in the case of deceased individuals, the specific extent of data protection rules after the death of individuals depends on national legislation. The GDPR has made this even more explicit. Recital 27 states: 'This Regulation does not apply to personal data of deceased persons. Member States may provide for rules regarding the processing of personal data of deceased persons."

[152] In other words, the Convention, depending on the legislation of a Council of Europe member state, was potentially applicable to legal persons. Collective entities without legal personality could also have been considered Data Subjects. The modernised Convention 108 of 2018 retained the broad notion of a Data Subject ('any individual'). However, the new Article 3 no longer refers to the possibility for Member States to extend protection to legal persons. The Data Protection Directive did not exclude legal persons from the definition of Data Subjects. Recital 24 of the Directive made it clear that 'legislation concerning the protection of legal persons with regard to the processing of data relating to them is not affected by this Directive'. However, there was no specific data protection legislation for legal persons at EU level. Consequently, Member States were free to regulate the protection of personal data of legal persons. This was also confirmed by the CJEU, which stated that nothing prevents Member States from extending the scope of national legislation implementing the provisions of the Directive to areas not included in the scope of the Directive, provided that no other provision of Community law (now EU law) precludes it. Only a few Member States have extended the notion of Interested Parties to legal persons: Italy, Austria and Luxembourg. However, the repeal of the DPD has reduced the Member States' room for manoeuvre in extending the scope of definitions and protections in the EU data protection framework, since the GDPR is a regulation requiring higher harmonisation. The preamble of the GDPR is more assertive on the issue of subjectivity of data of legal persons than the DPD. Recital 14 states, "This Regulation does not cover the processing of personal data relating to legal persons and in particular to companies incorporated as legal persons, including the name and form of the legal person and the contact details of the legal person."

27)[153] , that Member States may provide for rules regarding the processing of personal data of deceased persons, there is no reference to Member States' derogations regarding the protection of personal data of legal persons. Consequently, considering that the Member States have no margin of appreciation beyond what is explicitly stated in an EU regulation, it can be concluded that the notion of data subject cannot include legal persons, not even at national level. Moreover, whereas the title of the Data Protection Directive mentioned only individuals, the GDPR explicitly mentions only natural persons. Regarding the third requirement, the meaning of 'identified or identifiable, directly or indirectly', it is Article 4 that clarifies the list of identifiers: 'a name, an identification number, location data, an online identifier or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'[154] . Furthermore, Recital 26[155] also explains what identifiable means: 'To determine whether a natural person is identifiable, account should be taken of all the means likely reasonably to be used, such as distinguishing, either by the controller or by another person to identify the natural person directly or indirectly'. The notion of 'reasonably usable' is the key to this definition. It is, however, a slippery concept for which Recital 26) further states: to determine whether the means are reasonably usable to identify the natural person, all objective factors should be taken into account, such as the costs and time required for identification, taking into account the technology available at the time of processing and technological developments[156] . Ultimately,

[153] Recital 27) of EU Regulation 2016/679

[154] The WP29 has generally adopted a broad notion of 'identified or identifiable'. "Identified" refers to a known or distinct person in a group, and "identifiable" is a person not yet identified, but whose identification is possible. A person is directly identified or identifiable most commonly by reference to a name, in combination with additional information if the name is not unique; a person is 'indirectly identifiable' through so-called unique combinations of non-unique identifiers that make it possible to identify the individual.

[155] Recital 26) of EU Regulation 2016/679

[156] The CJEU has accepted this broad notion of identifiability (and the reasonable likelihood of identification) in the Breyer case(CJEU, C-582/14, 19/10/2016), where the Court states that it is not necessary "that all information enabling identification ... must be in the hands of a single person'. More recently, the Nowak judgment - reaffirmed these concepts. The Court, in Breyer, also states that identifiability is not 'reasonably likely' if it is 'prohibited by law or practically impossible' due to a 'disproportionate effort in terms of time, cost and labour, such that the risk of identification actually appears insignificant'. In other words, the Court introduces the legality factor among the means of identification. The Court recently confirmed these considerations.

However, this issue is still very controversial: scholars have pointed out the contradictions between the risk-based definition of 'identifiability' in Recital 26 of the GDPR and the authoritative WP29 guidelines, whose text (referring to 'irreversible' anonymisation and 'reasonably impossible' re-identification) seems to refer to a 'zero-risk' approach.

In general terms, taking into account existing guidelines, CJEU judgments and academic discussion on the point, the resulting standard of reasonable likelihood of identification is rather broad and context-dependent, the consequence of this broad definition being that the status of data as 'personal' is dynamic, i.e. the same dataset might not be identifiable (referring to an identifiable Data Subject) at the beginning of the processing from the perspective of a particular data controller, but might become identifiable (or start to be) from the perspective of another data controller or once circumstances change.

it can be said that the nature of the 'data subject' is dynamic, just as personal data are dynamic, changing over time. The relational, risk-based and context-dependent definition of the data subject is something, however, to be explored in depth, in particular with regard to the notion of data subjects and vulnerable subjects. Regarding the fourth requirement, the definition of the meaning of 'concerning' (an identified or identifiable individual) is also very broad. The WP29 interpreted 'concerning' as referring to three possible cases: the relationship could be based on content, i.e. the information would concern a natural person; on purpose, i.e. the data 'are used or are likely to be used for the purpose of assessing, treating in a certain way or influencing the status or behaviour of an individual'; or on impact, as the information, regardless of its content or any processing purpose, may concern an individual because its use is likely to affect the rights and interests of a particular person[157] . To sum up, the meaning of 'concerning' in the EU definition of personal data-and thus in the definition of data subject-is broad and dynamic: if the data processing concerns an identified or identifiable person or even if it is only potentially intended to deal with or have an impact on a particular identifiable person, the information processed is personal data and that person is a data subject. Understanding the relationship between the data subject and the data controller is relevant because it helps to qualify the level of vulnerability of a Data Subject. Indeed, the notion of a vulnerable (or non-vulnerable) data subject is relational: the encounter between the personal characteristics of a given data subject and the position of power of a given data controller characterises the data subject's level of vulnerability. The subject under consideration is substantially different from the normative notion of an informed and rational individual: in fact, individuals may not only not be fully aware of their rights and of

This is particularly true if one also considers Article 11, according to which, if the purposes for which a data controller processes personal data do not (or no longer) require the identification of a Data Subject by the data controller, that data controller is not obliged to take steps to identify the Data Subject for the sole purpose of complying with the GDPR. Moreover, in such cases, if "the data controller is able to demonstrate that it is unable to identify the Data Subject, the data controller shall inform the Data Subject accordingly" and "Articles 15 to 20 shall not apply unless the Data Subject, for the purposes of exercising his or her rights under those Articles, provides additional information enabling his or her identification". This provision seems to add more complexity not only to the difference between anonymous and personal data (including pseudonymised data), but also to the very definition of the data subject. Indeed, if we agree that under Article 4(1) a person who is not (or is no longer) identifiable cannot be called a 'Data Subject', we might wonder why Article 11, in both paragraphs (1) and (2), continues to refer to such persons as 'Data Subjects'. Article 11 suggests that non-identifiable Data Subjects (in particular individuals to whom anonymous data could be referred to if they provide further information enabling their identification) could exercise data protection rights. Apart from this grey area of subjects-which could be seen as occupying an intermediate zone between Data Subjects and non-Data Subjects-Article 11 reveals that the correct application of the principles of purpose limitation and data retention limitation affects the identifiability of the subject: when the purpose ends, identifiability should end; when the purpose does not require identifiability at all, such identifiability should not be made possible.

[157] The CJEU initially interpreted 'concerning' only as content-based (information about a natural person). More recently, in the Nowak case (CJEU, C-434/16, 20/12/2017), the Court adopted a similar approach to WP29, accepting that the relation can be by content, purpose or effect.

the risks involved in data processing, but in many circumstances they may also be inclined to give their consent without an adequate assessment of the implications, easily accepting any request. In the context of EU consumer law, distinguishing between average and vulnerable consumers, it is possible to examine whether the notion of the average consumer can be equated with a standard or average data subject. The rational and informed understanding typical of average consumers seems to correspond to the notion of data subject under the GDPR. Furthermore, the average data subject could be compared to a rational, circumspect and informed person. However, the Court of Justice of the European Union has recently adopted a more flexible view, applicable to specific contexts. Nevertheless, the growing awareness of cognitive biases also in 'average' individuals and empirical studies on the expectations and behaviour of data subjects have shown that such an optimistic view of the average data subject is unrealistic and potentially counterproductive. Consequently, it appears necessary to examine the notion of the vulnerable data subject in greater depth, starting with a theoretical reflection on vulnerable individuals[158] . The concept of vulnerability has been discussed in numerous fields, including political philosophy, gender studies, law, ethics, sociology, and others. Scholars from these different disciplines frequently confront each other, adapting conceptions of vulnerability developed in different contexts to enrich and broaden the understanding of the phenomenon in each field. In spite of the tension between universal vulnerability (everyone is vulnerable) and particular vulnerability (some people are more vulnerable than others), a common point for vulnerable individuals seems to lie in the imbalance of power and the increased risks to fundamental rights and freedoms arising from a structural, institutional or relational imbalance of power. If we focus on EU law, despite the existence of specific lists of vulnerable individuals in various sectors, the overall picture reflects a highly contextual and relational understanding of vulnerability, based on power imbalance and higher risks of harm. Vulnerability can, therefore, be defined as the significant probability of incurring identifiable harm without being able to protect oneself. The link between vulnerability and harm, in fact, recurs in many philosophical-legal reflections and legal and economic analyses, including those related to consumer vulnerability and clinical research. However, it is crucial to clarify that the concept of 'harm' can vary significantly between different contexts, legal and non-legal. Vulnerability, however, is not only related to demonstrable harms, but above all to risks of interference with fundamental rights and freedoms. Furthermore, the analysis of vulnerability should consider not only the personal

[158] On this point, see more extensively Malgieri, G. (2023). Vulnerability and Data Protection Law. Oxford University Press, Chapter 2.

characteristics of the individual, but also the interaction between these characteristics, the power of the counterpart and the social or technological mechanisms involved. The 'layered' vulnerability theory emphasises that these layers are not fixed attributes, but characteristics constructed by status, time and location. The assessment of layers of vulnerability should be based on an analysis of the origins and effects of this vulnerability . [159]

2. Digital vulnerability in European private law.

The topic of digital vulnerability and legal responses to it is a complex issue involving different spheres of law, requiring careful reflection on how jurisprudence, legislation and doctrine should address it. In this context, it is possible to outline some general aspects that deserve to be studied in depth, such as the importance of digital vulnerability for European private law, its nature as a cross-cutting social and legal problem, and the different legal responses, ranging from the use of traditional protection instruments to the introduction of new concepts. The project of European integration, centred initially on the internal market and later on the 'digital single market', has always aimed at combining economic orientation with the protection of weaker subjects, affected by structural disadvantages or specific vulnerabilities. This approach has found expression in policies aiming at protection against gender or racial discrimination, social protection of workers and consumers, and the safeguarding of fundamental rights. The balance between market development and protection of the vulnerable has characterised the evolution of European policies, as demonstrated by the directives on non-discrimination, labour law and consumer law. With the transition to the digital era, it has become necessary not only to perfect the digital single market, but also to adopt new protection policies, designed to counter the risks arising from digitisation, including those related to digital vulnerability. In this context, the legal challenges facing the European Union are not limited to the public dimension, but also involve private law. Although initially the focus of EU legislative action was mainly on public law, since the 1960s it has progressively extended to private law issues as well, with the creation of a true 'European private law'. This term describes the set of rules that the Union has developed to regulate transnational private matters, in a context where the national law of the Member States remains relevant. European private law is today a crucial element for the functioning

[159] On this point, see more extensively Malgieri, G. (2023). Vulnerability and Data Protection Law. Oxford University Press, Chapters 2 and 3.

of the internal market and protection policies, as evidenced by numerous regulations on discrimination, labour law, SME protection, and consumer protection. European private law is not, therefore, the exclusive result of digital dynamics, but has deeper roots going back to the private codifications of the 19th century, inspired by the liberalism prevailing at that time. However, the challenges posed by digitalisation call for a review of the basis and mechanisms for applying these principles. European Union legislation has already taken significant steps in this direction, with legal acts concerning consumer protection in the digital sphere, such as the 2019 'twin directives'[160] , which regulate contracts for digital content and services, as well as purchases that include goods with digital elements. Similarly, the Digital Services Act introduces provisions that aim to protect consumers and other fundamental rights in the online environment, while the adaptation of producer liability to technological developments also stems from increasing digitisation. The examples mentioned above show that the European Union is addressing new legislative challenges through both main modes of its regulatory intervention: on the one hand by harmonising the laws of the Member States through directives, and on the other hand by creating a uniform European law that is directly applicable through regulations. In addition to this recent regulatory development, which is mainly part of the European Commission's 2015 'Strategy for the Digital Single Market'[161] , there is now the question of whether and, if so, how the emerging phenomenon of digital vulnerability requires further development, both conceptually and legislatively, of European private law in relation to this challenge. Regarding the difficulties associated with the term 'digital vulnerability', one has to consider a variety of areas of everyday life and, consequently, a variety of legal issues. This consideration is even more relevant if one takes into account the ever-increasing impact of artificial intelligence on society and the everyday lives of individuals, an impact that will continue to expand in the future. The extent and diversity of the effects of digitisation make it impossible to reduce 'digital vulnerability' to a single type of risk or a specific harmful action. Rather, research in this area must address a plurality of forms and causes of vulnerability, as well as different conceptions of 'vulnerability' in the digital context. Despite this variety, a common feature of the phenomena encapsulated under the term 'digital vulnerability' is probably the fact that the use of digital technologies contributes to creating or increasing the risk of harm to the rights or interests of individuals or groups, in one of the multiple contexts of digital communication or enjoyment of digital content where such risks may manifest themselves. A significant example of this is the wide

[160] Reference is made to D irective (EU) 770/2019 and Directive (EU) 771/2019.
[161] https://eur-lex.europa.eu/IT/legal-content/summary/the-eu-s-new-digital-single-market-strategy.html

range of dangers associated with the use of digital devices that, by allowing the collection of personal data, expose the user to the risk of violation of their privacy or unauthorised access to such information. In this scenario, digital vulnerability issues are intertwined with data protection concerns and the regulation of data commerce, from both public and private law perspectives. Furthermore, the need for protection arising from digitisation may arise in situations where a user is dependent on the use of artificial intelligence and the processing of information beyond his or her control. Such dependence may generate a number of risks that are specifically related to digitisation, particularly in its advanced application through artificial intelligence. In some cases, however, it may be difficult to distinguish between 'digital risks' and pre-existing threats that do not depend on a digital context, as well as to establish whether the digital environment has increased or reduced vulnerability. For instance, where consumers use artificial intelligence to conclude online purchase contracts, such use may offset some of the disadvantages of information asymmetry with respect to the seller. On the other hand, however, they become dependent on the artificial intelligence provider, which in certain situations may also act on behalf of both parties to the contract (as happens in *machine-to-machine* purchases). In such circumstances, the digitisation of the contract could, depending on the specific case, either improve the position of consumers or increase the risks related to their vulnerability, making a definitive assessment complex. For these reasons, in-depth reflection will be needed to adequately classify the vulnerability in these cases and to further develop regulations to protect consumers who conclude contracts on the basis of these technologies. However, it seems questionable to apply the concept of digital vulnerability to all situations in which new risks, related to both tangible and intangible assets, arise from the use of digital technologies, and to assume that those involved necessarily need legal protection in such contexts. Indeed, it is necessary to examine the extent to which the parties concerned are, in principle, able to protect their own interests, also in relation to these new circumstances, within the framework of private autonomy. If this capacity for self-protection proves to be lacking - for example, due to a structural imbalance between the parties or due to situations of inequality in the legal relationship that make it difficult for one party to defend its interests - legal measures may be necessary to protect against the consequences of digitisation. Examples of such measures include general provisions such as those in the GDPR for the protection of personal data or those relating to the liability of software producers in the context of the Product Liability Directive. In this case, the situations in question could be regarded as 'digital vulnerability' in a broad sense. However, it remains to be seen whether a narrower interpretation of the term would be more

appropriate than such a broad use of the concept. A narrower understanding of 'digital vulnerability' might be particularly appropriate, especially in light of terminology already developed in consumer law, such as 'vulnerable consumer' (as opposed to 'average consumer'). In such an approach, 'digital vulnerability' would not include all risks arising from digitalisation that could affect any subject in the legal relationship, but would rather focus on those risks that emerge as a result of digitalisation and that affect particular groups of people, due to specific circumstances, such as their social or professional position or the stage of life they are in. Even in a narrower conception, however, the term 'digital vulnerability' would cover a wide range of very different situations, which highlight how such vulnerability is a cross-cutting issue, both in terms of the factual circumstances to be considered and the legal issues to be addressed. In fact, from a factual point of view, one has to consider in particular the social circumstances that, through digitalisation, may lead to a specific vulnerability for certain groups of people. These include, for instance, aspects related to education, income and distribution of social roles, which might negatively affect participation in digital communication, the ability to use information and the protection of one's interests in the digital context. No less relevant, in terms of factual circumstances, is the fact that digital vulnerability may result from physical impairments or, more frequently, from issues related to particular stages of life. In this sense, digital vulnerability extends to a broad spectrum of situations, from 'child vulnerability' (as in the case of 'smart toys' for children or digital manipulation and forms of online bullying at school) to the exploitation of the 'digital weaknesses' of the elderly. These areas of digital vulnerability clearly show that risks are not only limited to material damage, but also include risks related to serious immaterial damage. Moreover, digital vulnerability does not end with death either. The 'eternal memory' of the Internet allows access to what individuals wished to keep private during their lifetime. More significantly, the manipulative power of artificial intelligence can alter the image that future generations have of a person, changing their representation in a literal and figurative sense, for instance through what they wrote or created. In other words, *post-mortem*, the real personality can be replaced by fiction, without the person, whose identity has been violated, being able to defend himself. The variety of situations that can give rise to digital vulnerability is reflected in the wide range of legal issues in which legislation and enforcement must respond to this challenge. In other words, the broad spectrum of digital vulnerability as a phenomenon affecting many areas of society in the digital age manifests itself at the legal level as a cross-cutting challenge for almost all branches of law, judicial sectors and legal disciplines. Similar to national law, it is therefore necessary for European private law to

address the multiple types of digital vulnerability within the various private legal relationships, taking into account the context of each branch of private law. These branches partially overlap with areas already covered by EU legislation in protection contexts, such as consumer law (including consumer insurance), tort law, unfair trade practices legislation, labour law regarding digital vulnerability of employees, online dispute resolution and private international law. In all these and other areas, research in private law is faced with the task of analysing the new problems arising from digitisation in relation to digital vulnerability and proposing appropriate solutions in the respective legal contexts. In contrast to such analyses in the context of individual legal fields, it would probably be less fruitful to try to completely isolate all digital vulnerability issues from their specific legal contexts and treat them as a new, separate legal field of 'digital vulnerability'. Indeed, such an approach would be as questionable as the idea that 'digital law' as a whole should be separated from 'analogue law' as an independent field (and perhaps that in the future lawyers will specialise in 'digital law' and 'analogue law', instead of private law, public law, etc.). It is not clear that such a separation and isolation of 'digital law' from the rest of the legal system can be justified by the current and foreseeable development of positive law at the European or national level. Moreover, it would not improve legal protection for digitally vulnerable groups, but would only lead to an unnecessary complication of the legal protection system. It is therefore considered more appropriate to consider, as far as possible, the specific context of each area of law in order to address the legal issues generated by digitisation in relation to digital vulnerability and, accordingly, to adapt the respective branches of private law to the new challenges. Although the search for legal answers to digital vulnerability is a relatively recent challenge, research contributions in the legal areas just mentioned already suggest a number of approaches to further develop European private law in this respect. These approaches concern both the adaptation of existing protection instruments to the changes resulting from digitisation and the development of new concepts specifically related to digital vulnerability. On the one hand, the potential of already existing protection instruments is being assessed. Examples of this are information obligations to compensate for information asymmetries on the Internet, withdrawal rights for online transactions or mandatory regulations on contractual compliance of digital products. On the other hand, it is also necessary to consider the possibilities of extending and supplementing these instruments, e.g. through the liability of digital intermediary services or the use of appropriate technologies such as 'filters' to prevent certain forms of interference. Digital vulnerability is a challenge on several levels in several respects. Not only in terms of the ways and levels at which it manifests itself, but also in terms of the

levels at which legal responses must be developed. In the latter case, one does not only consider the various levels within a legal system: the constitution, private law regulations, law enforcement by the courts, alternative dispute resolution, self-regulation through codes of conduct, etc. Rather, special attention should be paid to the fact that the digital revolution, and the associated digital vulnerability, poses a multi-layered challenge precisely because of its cross-border nature, especially in the relationship between the national law of one state, the supranational law of the European Union and the international law beyond. In this context, EU law deserves special attention as an intermediate level. This focus on European law reflects, on the one hand, the cross-border nature of digitisation and its relevance for the European single market. On the other, it is due to the fact that, although global approaches to legal solutions across European borders are desirable, they seem to be only possible to a very limited extent at present, mainly due to political differences between the home states of the 'global players' of digitisation, including divergent approaches to assessing the consequences of digitisation. Indeed, it seems that Europe is charting its own course with respect to legal responses to the implications of digitisation, contrasting not only with China, but also with the US in various areas, from data protection (as regulated in the GDPR) to market regulation (through regulations such as the DMA, the DSA and the Data Act[162]) and liability for digital products (as dealt with in the new proposal for a directive on product liability and the Commission's proposal for a directive on liability for AI). However, even from this European perspective, it is necessary to consider the legal framework and possibilities for action that extend beyond the borders of Europe. In this context, a number of projects aimed at adapting international regulations to technological changes and the resulting new protection requirements deserve attention. In view of the cross-border nature of digital transactions, the question of how digital vulnerability can be integrated into the development of private international law becomes particularly relevant. This issue will also have to be taken into account in international legal unification efforts, in its many forms, from conventions to model laws. An example of such initiatives is the UNIDROIT projects on new technologies in the digital age, and the work of UNCITRAL, which is exploring the adaptation of international sales law to the changes brought about by digitalisation. In the past, the Vienna Convention on the International Sale of Goods of 1980 had a lasting impact on European contract law. With regard to the renewed modernisation of contract law, which appears to be necessary due to digitalisation at both European and international level, there could be a mutual stimulation of projects on both levels. Innovations already introduced in

[162] Reference is made to Regulation (EU) 2022/1925; Regulation (EU) 2022/2065; Regulation (EU) 2023/2854.

EU legislation, such as the digital content directive regarding the updating of obligations and legal consequences of termination, could inspire the development of international sales law. In addition, considerations on automated contracting, relating to international trade law, could also influence European legislation in this field, resulting in a fruitful dialogue between the various legal levels. A significant feature of many of the efforts to find legal answers to digital vulnerability is the interplay between private and public law approaches. This includes the legal basis for protection against infringement, considering that rights such as honour, other personal rights and individual property rights are traditionally protected by private law. However, these rights are also protected by fundamental rights, in particular at European level, by the Charter of Fundamental Rights of the European Union. The Charter of Fundamental Rights of the European Union is of particular importance with regard to digital vulnerability, as it establishes the right to the protection of personal data (Art. 8) and consumer protection (Art. 38)[163] . Based on these principles, the EU has combined provisions relevant to digital vulnerability in both public and private law, sometimes even in a single set of rules. An example of this is the protection of personal data, which is mainly regulated by public law in the General Data Protection Regulation (GDPR), and the purchase of digital products, which is mainly dealt with by private law in the Digital Content Directive. The analysis and further development of legal responses for the various types of digital infringements must, therefore, take into account the tension between, on the one hand, private law-based approaches and, on the other hand, the fundamental values protected, such as the protection of private autonomy in contract law and the fundamental rights enshrined in the Constitution. Consequently, questions arise as to the preference of public law measures, such as bans on certain online content, over private law instruments, such as the liability of users or intermediary services, depending on the specific circumstances. In these issues, the analysis of the causes and forms of digital vulnerability and the development of legal responses cannot be limited to the boundaries of private law, but must also consider public law perspectives. Another relevant issue concerns the relationship between consumer protection and the protection of Internet users. Not only are the challenges of protecting against digital vulnerability largely consumer protection issues, but the question also arises as to whether the protection of Internet users should be extended, or whether a new and broader concept of user protection should emerge in place of the traditional concept of consumer protection. Recent research highlights the first aspect, exploring requirements and approaches for the development of European consumer law. This includes general

[163] Art. 38 of the Charter of Fundamental Rights of the European Union.

reflections on digital vulnerability in EU consumer law and a number of specific aspects, such as the gap between digital fairness and transparency, and the digital vulnerability of insurance consumers in relation to personalised pricing of insurance products. Similarly, the evolution of European private international law is being analysed in relation to the 'digital vulnerability of consumers in the world of smart contracts'. However, the traditional concept of consumer protection does not provide a sufficient framework to fully address the problem of digital vulnerability and to develop comprehensive legal responses. The current discussions on digital vulnerability and related legal protection measures are therefore at a point of tension between the traditional concept of consumer protection and the focus on the protection of Internet users, or at least users of certain forms of digital communication. The issue of user, as well as consumer, protection is particularly important since Internet users are exposed to digital vulnerability in situations where they are not protected as consumers according to the traditional understanding of EU law, such as in the case of infringements caused by actors other than traders or by traders with whom they do not have a legal relationship or do not wish to establish one. In these situations, it is likely to be difficult to adequately develop the protection of data subjects from risks related to new technological developments based on the current conception of the consumer in EU law. Therefore, it is necessary to reconsider whether, and to what extent, in the light of these developments, an alternative concept, such as 'user', should replace or at least complement the traditional role of the consumer. This question concerning the relationship between the concepts of consumer and user emerges, for instance, in reflections focusing on the issue of digital vulnerability in the 'post-consumerist society', as well as in critical analyses of personalised pricing and the notion of consumer in relation to traditional sectors of protection, such as consumer insurance[164] . From the framework just outlined, it emerges how the challenges posed by profiling and automated processing may persist, even if they have already been partially resolved and regulated, with reference to various situations in which the vulnerable person is a protagonist.

3.Limitations and alternatives to the *vulnerability-based* approach

The assessment of layers of vulnerability through the Data Protection Impact Assessment (DPIA) model has enabled the development of an interpretation of the General Data

[164] For the full paragraph see Schulze, R. Digital Vulnerability in European Private Law - Conclusions, forthcoming.

Protection Regulation (GDPR) that takes vulnerability into account. However, this approach is not without criticism, and there are several potential solutions to the problems it presents, as well as alternative models to consider. Possible disadvantages of an inclusive approach to the vulnerability of Data Subjects and the consequent vulnerability-sensitive interpretation of the GDPR include (i) the limitation of the autonomy of Data Subjects, which could result in paternalism; (ii) the lack of legal certainty and predictability; (iii) the risk of an overly individualistic approach; (iv) the limitations of analogical interpretation; and (v) the inconvenience and ineffectiveness of an extensive interpretation of several provisions of the GDPR[165] . Regarding the first point, one of the main criticisms that can be made of a vulnerability-aware approach to data protection is the risk of limiting the autonomy of the data subject. Some scholars[166] , analysing both US and EU legislation in 2013, have criticised the 'paternalism' inherent in the default rules of data protection law, which conflict with the consent of Data Subjects. These scholars argue that choices regarding the processing of one's own data are not always clear to individuals, who may not care about the processing of their data and may be willing to accept being profiled, targeted or tracked, if this entails benefits such as free online services or optimised personalisation of content. According to Solove, the cost-benefit analysis does not unequivocally recommend not disclosing data, as the data subject is in the best position to weigh up the risks and benefits. Moreover, privacy is inherently contextual, which implies that decisions on the management of personal data should be tailored to the specific circumstances of individuals. While recognising that self-management of personal data, based on consent and individual rights, may have serious structural and cognitive shortcomings, many authors believe that the only individual who can clarify his or her expectations in a given context is the data subject, through consent and related rights. In this way, the subject reveals what he or she considers appropriate in a specific situation. On the contrary, according to some scholars, limiting the data subject's opportunities to choose how to process their data, by reinforcing the data controller's duties of responsibility, would benefit precisely the data controllers, who could interpret data protection rules in their own favour. Similarly, other doctrine[167] argues that shifting data protection from individual consent and rights to the data controllers' duties of responsibility

[165] Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press, p.189

[166] Solove, D. J. (1880). 'Privacy Self-Management and the Consent Dilemma' (2013). Harvard Law Review, 126, 1880 reported by Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press. P. 191.

[167] Cavoukian, A., Dix, A., & El Emam, K. (2014). The unintended consequences of privacy paternalism. Information and Privacy Commissioner of Ontario, Canada as reported by Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press. P. 191.

is excessively paternalistic, as it weakens the data subject's self-determination, participation and control in the field of data protection. According to this perspective, the shift from a model based on the autonomy of the data subject to a model centred on the responsibilities of data controllers would reduce the ability of individuals to make informed decisions with regard to their personal data, limiting their freedom of choice and their effective control over their information. Any reduction of the role of consent and individual rights, according to others[168] , threatens one of the legal foundations of the data protection framework: informational self-determination. Restricting consent in order to strengthen accountability obligations could shift the entire burden of verifying data protection compliance onto regulators and supervisory authorities, such as DPAs, who would be called upon to take on the difficult task of proving damage to data protection. Comprehensive protection could then be shifted to an *ex post* approach, where the harm-based model would only protect individuals after the harm has actually occurred. This approach is even more problematic considering the lack of regulatory clarity regarding the notion of harm in the context of data protection[169] . As to the second point, however, a further limitation in adopting a vulnerability-centred interpretation of data protection law lies in the difficulty for the data controller to anticipate with sufficient precision whether certain categories of vulnerable individuals will be affected by specific data processing operations, as well as to identify the vulnerable individuals in question. This uncertainty creates a problem of legal stability and predictability for data controllers, who find themselves unable to take adequate measures to implement data protection regulations. In other words, the adoption of the notion of vulnerability as a layered, damage-based concept requires an ex post assessment that cannot be carried out before processing operations begin[170] . A further argument critical of this approach to vulnerability lies in its potential individualism, which may overlook the reality that many forms of vulnerability are structural and collective in nature. In particular, homogeneous groups of vulnerable individuals may find themselves exposed to similar risks to their fundamental rights and freedoms. One thinks, for instance, of discrimination based on race, gender, sexual orientation, cognitive impairment related to age, mental disorders or socio-economic factors (such as nationality, language spoken or economic status), or

[168] Purtova, N. (2014). Default entitlements in personal data in the proposed Regulation: Informational self-determination off the table... and back on again?. Computer Law & Security Review, 30(1), 6-24 reported by Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press. P. 191.
[169] On this point, see more extensively, Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press. Pp. 190-196.

[170] On this point, see Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press, pp. 197-202.

limitations of freedom resulting from specific hierarchical structures (e.g. workers, soldiers, prisoners, etc.). Addressing the vulnerability of such individuals from an exclusively individual and case-by-case perspective may risk ignoring the underlying structural and social issues, such as social injustices, racial prejudices and other forms of discrimination, which are the primary cause of many of the vulnerabilities found[171] . A further significant problem with this proposal concerns legal interpretation. Indeed, in the text of the GDPR, the only category explicitly recognised as vulnerable is that of children. Applying by analogy the safeguards provided for children to vulnerable adults could lead to an overestimation of the importance of the category of 'vulnerable data subjects' as outlined in the regulation. Indeed, the vulnerability of individuals is mentioned explicitly only in the preamble, precisely in recital 75)[172] , without any real explication in the body of the normative text[173] . Another issue and limitation of a *vulnerability-based* approach could be the broad interpretation of GDPR .[174]

[171] On this point, consider Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press, pp. 202-205.
[172] Recital 75) GDPR .
[173] In this regard, see Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press, pp. 205-208. According to the author, the third limitation considered raises two main issues: (i) the technical limitations related to analogical interpretation and (ii) the risk of attributing undue importance to a concept that remains, in the GDPR, mostly implicit and not formally developed.
[174] On this point, consider what is advocated by Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press, pp. 209- 218.

Chapter III

'The limits and opportunities for regulation of profiling in relation to the vulnerable individual. *Large Language models* in the context of the AI Act'.

SUMMARY: 1. Artificial Intelligence: foundations, functioning, learning models. *Foundation Models* and *Large Language Models*. 2. The European Union Regulation on Artificial Intelligence: a new approach to digital? 3. *Large Language Models*. A test bench for the European legislator. 3.1. *Foundation Models* in the proposed Artificial Intelligence Regulation. 3.2. The text of the current Regulation: *General-purpose AI Models* and *General-purpose AI Systems*. 4. *Large language models,* profiling, vulnerabilities. 5. Perspectives *de iure condendo*: safeguards, *vexatae quaestiones* and opportunities: a "systemic vulnerability"?

1. Artificial Intelligence: foundations, functioning, learning models.

Since antiquity, several fundamental methodological questions in the philosophical field have been similar to those that arise today with regard to Artificial Intelligence. Questions such as "is it possible to automate reasoning?", "what are the basic cognitive operations?", "what are the necessary conditions for a formal language to be adequate to describe reality?" have been asked over the millennia of philosophical thought by philosophers such as Aristotle, St. Thomas Aquinas, William of Ockham, René Descartes, Thomas Hobbes and Gottfried W. Leibniz. However, the question "Is it possible to build an artificial intelligence system?" could only be answered by experiments conducted since the 20th century, when the first computers were built. In the field of Artificial Intelligence, in fact, research investigating the intelligence of a system built by a human designer plays a major role. On the basis of such reflections, Alan M. Turing proposed a solution to this problem in 1950 with the help of the so-called *imitation game* . [175]

Conventionally, 1956 is considered the year of the birth of Artificial Intelligence, as the lecture at Dartmouth College took place[176] . Another year of primary importance is 1959,

[175] In particular, the experiment performed by A. M. Turing consists of an operational test of Artificial Intelligence. It is to be assumed that a human interrogator simultaneously has a conversation with both another human being and a computer. This conversation is carried out with the aid of a device that makes it impossible to simply identify the interlocutor. The human interrogator, after some time, would have to guess which statements are sent by the human being and which are sent by the computer. According to Turing, if the interrogator cannot make such a distinction, then the (artificial) intelligence of the computer is the same as the intelligence of the human being. At this point, it is easy to see that intelligence is, to some extent, considered as linguistic competence.

[176] It is worth noting, in this regard, that not all scholars agree in identifying 1956 as the year of the birth of Artificial Intelligence. In fact, as early as 1955 the first artificial intelligence system, called Logic Theorist, was designed by Allen Newell, Herbert A. Simon, and implemented by J. Clifford Shaw at Carnegie Mellon University. The system proved nearly forty theorems included in the seminal monograph Principia Mathematica

when the *General Problem Solver*, GPS[177] , was implemented. The implementers themselves defined the *cognitive simulation* paradigm as the methodological outcome of their research. This paradigm states that general patterns of human ways of solving problems should be simulated in artificial intelligence systems. Among other things, it then becomes a basic assumption of advanced research projects on cognitive architectures. While Newell and Simon were building their systems on the basis of cognitive simulation, John McCarthy was conducting research at MIT in another crucial area of Artificial Intelligence. In 1958, he presented his important paper at the Symposium on the Mechanisation of Mental Processes, organised at Teddington. McCarthy proposed the paradigm of solving common sense problems with the use of reasoning models based on formal logic. At the time, such an idea seemed peculiar[178] . His research was successful and in 1958-1960, he built the Lisp language[179] . In the early 1970s, within the logic-based paradigm, a new research approach appeared based on first-order logic, FOL. The result was the construction of the second classical language of artificial intelligence, namely Prolog, by Alain Colmerauer and Philippe Roussel (University of Aix-Marseille II). This approach gave rise to the popular model of logic programming with constraints, CLP. In the mid-1960s, the third approach to symbolic AI appeared, namely the knowledge-based approach. It originated from a project to implement the Dendral expert system conducted by Edward Feigenbaum and Joshua Lederberg at Stanford University. The goal of the system was the identification of unknown organic molecules based on the analysis of their mass spectra and knowledge of chemistry, which proved very useful for organic chemists. Feigenbaum's experience led him to introduce a new paradigm of artificial intelligence that differs from both cognitive simulation and the logic-based approach. This paradigm can be described as follows. Firstly, instead of solving general problems, intelligent systems should focus on well-defined application areas[180] . Secondly, an intelligent system should be equipped with all the knowledge that human experts possess in a particular field:

---

by Alfred N. Whitehead and Bertrand Russell. The system's designers attempted to publish their results in the prestigious Journal of Symbolic Logic. The editors rejected the article, claiming that it only contained new demonstrations of elementary theorems and neglected the fact that a computer system was a co-author.

[177] This was also the result of Simon, Newell and Shaw's research into the construction of systems with mental abilities. In particular, the system they created solved a variety of formal problems, for example: symbolic integration, finding paths in Euler's problem of the Königsberg bridges, playing the Hanoi Towers puzzle and other issues.

[178] During a discussion, Yehoshua Bar-Hillel argued that inference based on logic (deductive) is not an adequate model for human reasoning based on common sense. This criticism did not deter McCarthy from the idea of using mathematical logic in artificial intelligence. He continued his research into a logic-based programming language for the implementation of intelligent systems. McCarthy was inspired by one of the modern logic calculi, the lambda calculus (λ-calculus), introduced by Alonzo Church and Stephen C. Kleene in the 1930s.

[179] Lisp and its dialects, such as Scheme and Common Lisp, are still used to build artificial intelligence systems.

[180] Such as the Dendral system.

therefore, such systems are often called expert systems. Thirdly, this knowledge should be treated as a kind of data and should be stored in the knowledge base of the system. A general inference mechanism should be implemented in the system to reason about the knowledge. The last assumption concerns a way of verifying the correctness of the system's functionality. The implemented system should be integrated into an environment where human experts solve problems. Its test is whether it simulates the experts well. If it does, it means that the system functions correctly[181] . These assumptions are largely fulfilled in the symbolic system building model of artificial intelligence and encouraged Newell and Simon to formulate a fundamental view of strong artificial intelligence in 1976, namely the hypothesis of the physical symbol system[182] . The hypothesis is formulated as follows:

A physical symbolic system possesses the necessary and sufficient means for general intelligent action.

This radical view of Newell and Simon that a system that transforms symbolic structures can be considered intelligent was strongly criticised by John R. Searle in his article "*Minds, Brains, and Programs*" published in 1980. Searle divided opinions on AI into two basic groups. Proponents of weak artificial intelligence treat the computer as a useful device for testing hypotheses concerning the brain and mental processes and for simulating brain performance. On the other hand, supporters of Strong Artificial Intelligence consider a properly programmed computer equivalent to the human brain and its mental activity. Of course, the hypothesis of the physical symbolic system belongs to the latter approach. To refute this hypothesis, Searle proposed a mental experiment, which also challenges the Turing test[183] . It can easily be seen that, starting with the Turing test, in Artificial Intelligence great significance is attached to natural language and the ability of intelligent systems to use it. Three linguistic theories, namely Noam Chomsky's theory of generative grammars, Roger

[181] In particular, this is analogous to cognitive simulation, where the system is supposed to simulate the intelligent behaviour of a human being.

[182] Specifically, a physical symbol system consists of a set of elements called symbols that are used by the system to construct symbolic structures called expressions and a set of processes for their modification, reproduction and destruction. In other words, the system transforms a certain set of expressions.

[183] One of the various versions of this experiment, called the Chinese room, can be described as follows. It is assumed that a native English speaker, who does not speak Chinese, is locked in a room. In the first phase of the experiment, he receives pieces of paper with Chinese characters on them and is asked to respond. However, he does not understand these notes. Fortunately, there is a book in the room that contains instructions on how to respond to the Chinese characters by writing other Chinese characters on a piece of paper. So each time he receives a piece of paper with Chinese characters on it, he 'produces' an answer according to these instructions. In the second phase of the experiment, he receives pieces of paper with sentences in English, e.g. some questions, and is asked to answer them. He then replies. It is then assumed that the quality of the answers in both cases is the same. Now Searle asks the question: is there any difference between the two phases of the experiment? In the scientist's opinion, the answer is positive. In the first phase the man in the room does not understand what he is doing, but in the second phase he does.

Schank's theory of conceptual dependency and George Lakoff's cognitive linguistics are particularly important in artificial intelligence. Noam Chomsky introduced the first version of generative grammars in 1957. According to his theory, the ability to learn a natural language grammar, called a universal grammar, is a property of the human brain[184] . This grammar takes the form of a formal system, which consists of a finite set of rules that are used to generate an infinite set of sentences belonging to the language. Grammar is thus the generator of language. Chomsky distinguishes not only a syntactic level but also a semantic level in the structure of a generative grammar. Furthermore, syntax constitutes a basic, primary level, while semantics is somehow derived from syntax. Thus, there are some similarities with the physical symbolic system model. Although the initial expectations of natural language processing with the help of this theory were not completely fulfilled, it has become a model of fundamental importance in computer science. The birth of this theory can be regarded as the origin of the dynamic development of mathematical linguistics. It focuses on the definition of mathematical formalisms that can be used to represent structures of both formal and natural languages. Roger Schank, in his theory of conceptual dependencies developed in the late 1960s, argued that it is not syntax that should be a starting point for defining the semantics of language, but concepts, strictly speaking dependencies, i.e. relations, between concepts. In order to formalise various types of dependencies between concepts, he defines a model for representing relations in the form of conceptual dependency graphs. These graphs are defined in such a way that two linguistic structures with the same meaning are represented by the same graph. In Artificial Intelligence, conceptual dependency theory constitutes a paradigm for structural models of knowledge representation. In these models, knowledge is represented by graph-like or hierarchical structures[185] . In the field of linguistics, cognitive linguistics as represented by George Lakoff seems to oppose the physical symbolic system hypothesis to a large extent. In his famous book 'Women, Fire and

[184] The universal grammar is 'parameterised' when a child is exposed to a specific language, which creates a grammar of that language in his or her brain.

[185] The semantic networks introduced by Allan M. Collins and Ross Quillian, the frames defined by Marvin Minsky and the scripts proposed by Schank and Robert P. Abelson are the most popular models in this approach. Research in this area initially focused on Natural Language Processing, NLP. The result was the construction of well-known NLP systems such as ELIZA simulating a psychotherapist, built by Joseph Weizenbaum in 1966, SHRDLU, the first system capable of understanding natural language in the context of a simplified world-like environment (TerryWinograd, 1970), MARGIE (Roger Schank, 1973) and SAM (Richard E. Cullingford, 1978), which were based on conceptual dependency theory, and PAM, built by Robert Wilensky in 1978, which was capable of interpreting simple stories. The approach based on structural models of knowledge representation has also been applied to build artificial intelligence systems beyond natural language processing. A variety of languages and programming environments were developed to implement such systems, including KRL defined by Terry Winograd and Daniel G. Bobrow and KL-ONE introduced by Ronald J. Brachman and James G. Schmolze.

Dangerous Things: Which Categories Reveal the Mind', published in 1987, he launched the most effective attack on strong AI. Instead of considering whether a machine that manipulates symbolic expressions can be called intelligent, G. Lakoff posed the question: "is it possible to define a mapping from our real-world environment into a set of such symbolic expressions?" He pointed out that somehow hidden in the hypothesis of the physical symbol system was an assumption about the possibility of the construction of such a mapping, which, however, was based on the classical Aristotelian approach to the creation of concepts. According to this approach, concepts have clear boundaries, i.e. an object either belongs to a category that refers to a concept or it does not. However, Lakoff argued that such an approach should not be taken in the light of modern psychology and philosophy. Ludwig Wittgenstein and Eleanor Rosch have shown that there are 'better' and 'worse' representative members of a category and that both human (bodily) experience and imagination play a fundamental role in the process of categorisation (conceptualisation). Therefore, categories should not be treated as 'boxes' that include some objects and not others, but should be defined taking into account their confusion. This confusion stems from the fact that our cognition is determined by the shape of our (human) body, in particular the shape of our brain (the thesis of the embodied mind). Therefore, cognitive linguists often argue that connectionist models, which mainly include models of artificial neural networks, are more appropriate for building artificial intelligence systems than methods based on logic and symbolic processing. Connectionist models, in which mental phenomena are modelled as emergent processes occurring in networks consisting of elementary components, have a long history in artificial intelligence. The first model of an artificial neuron that could be used as an elementary component of such a network was defined by Warren S. McCulloch and Walter Pitts in 1943. Neural networks, NN, are used to simulate the processes that occur in biological neural networks in a brain. They consist of simple interconnected components, which process information in parallel. From a functional point of view, a neural network is a kind of 'black box' that can be trained to learn appropriate responses to various stimuli. This learning consists of changing the parameters of a neural network structure. Consequently, this knowledge is in some way encoded in such a parameterised structure. Therefore, a connectionist approach differs completely from a symbolist approach. Therefore, one of the most debated questions on Artificial Intelligence took place between the proponents of these two approaches. In 1957, Frank Rosenblatt defined the perceptron, which at that time was a simple one-layer neural network. Subsequently, Marvin Minsky and Seymour Papert published their famous book 'Perceptrons' in 1969, in which they showed

the strong limitations of perceptrons, e.g. the inability to calculate certain logical functions such as XOR. As a result, many researchers in the field of Artificial Intelligence have come to the conclusion that the study of neural networks is not promising. Although such negative opinions from eminent authorities in the field of artificial intelligence caused limited funding of neural network research, it continued in the 1970s and 1980s and produced many positive results. In 1972, Teuvo Kohonen built the association network. Three years later, Kunihiko Fukushima, a senior researcher at the scientific and technical research laboratories of the Japan Broadcasting Corporation, built the *Cognitron*, which was a multilayer neural network. In 1982, two important models were developed: Hopfield's recurrent network and Kohonen's Self-Organising Maps, SOM. David E. Rumelhart and Geoffrey E. Hinton and their collaborators published a paper on the learning of multilayer networks using the backpropagation method in 1986. The following year, Stephen Grossberg and Gail Carpenter defined neural networks based on adaptive resonance theory, ART. After a series of successful research results, the return of the connectionist model and neural networks came in 1987, when the First International Conference on Neural Networks was organised. At the conference, its chairman Bart Kosko announced the victory of the neural network paradigm. Proponents of artificial neural networks assume that simulating nature at its biological level is an appropriate methodological principle for building artificial intelligence systems. They have attempted to simulate the brain in both its anatomical/structural and physiological/functional aspects. A simulation of nature in its evolutionary aspect is a basic paradigm for establishing biologically inspired models, which include an important group of artificial intelligence methods called evolutionary computation. These methods are used in the crucial area of finding an optimal solution to a problem. The basic principles of this approach include generating and analysing many potential solutions in parallel, treating a set of these potential solutions as a population of individuals, and using operations on these individual solutions that are analogous to the operations of biological evolution such as crossover, mutation, and natural selection. Alex Fraser published the first article presenting the ideas of evolutionary computing in 1957; therefore, he can be considered the pioneer of this paradigm[186] . Methods built on the basis of various mathematical theories have played a key role in the development of Artificial Intelligence models from the very beginning. The

[186] Within this approach, four basic groups of methods were developed: genetic algorithms, which became popular when John Holland published his monograph in 1975; evolutionary strategies, which were introduced in the 1960s by Ingorechnerberg and Hans-Paul Sulfur; evolutionary programming, devised by Lawrence J. Fogel at the University of California, Los Angeles in the 1960s; and genetic programming, popularised by John Koza in the 1990s.

most important ones include pattern recognition, cluster analysis, Bayesian inference and Bayesian networks, models based on fuzzy set theory and models based on approximate set theory. Pattern recognition is one of the oldest fields of artificial intelligence. In fact, it originated as an area of research before the term Artificial Intelligence was coined, because the first method within this approach was developed by Ronald A. Fisher in 1936. Pattern recognition methods are used for the classification of unknown objects/phenomena, which are represented by sets of features. The complementary issue of cluster analysis is to group a set of objects/phenomena into classes (categories). Bayesian theory was used very early on in artificial intelligence. Its role in artificial intelligence increased even more when Judea Pearl introduced an inference model based on Bayes' networks. The imperfection of a system of notions used to describe the real world is a crucial problem if we apply mathematical models of reasoning. Since the world is complex and multi-aspect, the concepts used in its representation are usually unambiguous and imprecise, which is unacceptable. To solve this dilemma, Lotfi A. Zadeh introduced fuzzy set theory in 1965 and in the 1980s Zdzisław Pawlak developed approximate set theory.

Artificial intelligence, since its inception - and despite the short period that did not see it play a leading role on the horizon of the scientific community in the strict sense of the term - has always been an interdisciplinary research area with peculiar and fascinating profiles of interest. It is, for these reasons, also a place of heated controversy between adherents of different schools. In principle, these disputes are philosophical in nature, as different approaches have been developed since the 1980s.

With the evolution of the state of the art in this field, and especially through the implementation of *machine learning* and *deep learning* techniques (see *below*), increasingly sophisticated Artificial Intelligence models have been achieved. This represented a challenge of unprecedented scope for the European legislator, but not only. In order to better understand the context in which the legislator found himself moving, it is necessary to briefly mention the learning mechanisms of Artificial Intelligence. In this regard, it seems of primary importance to deal with so-called *machine learning*. This term is used to identify a technology that enables machines to perform various tasks without having been explicitly programmed for the purpose. Fundamental to this is the availability of a vast set of data, the *dataset*. In other words, *machine learning* is the technique that improves system performance by learning from experience using computational methods. In computer systems, experience exists in the form of data and the main task of *machine* learning is to develop learning algorithms that build models from the data. By feeding the learning algorithm with experience data, a model

is obtained that can make predictions about new observations[187] . In summary, the term *machine learning* refers to the automated detection of meaningful patterns in data. In recent decades, it has become a common tool in almost all activities that require the extraction of information from large data sets. Furthermore, a common feature of the various applications of *machine learning* is that, unlike more traditional uses of computers, in these cases, due to the complexity of the patterns that need to be detected, a human programmer cannot provide an explicit and detailed specification of the parameters how such tasks should be performed. Taking the example of intelligent beings, many of the skills of human beings are acquired or perfected by learning from human experience (rather than by following explicit instructions that are provided). Machine learning tools are concerned with equipping programmes with the ability to 'learn' and adapt[188] . On the basis of these observations, in order to understand the scope of *machine learning*, it seems appropriate to dwell preliminarily on the concept and dynamics of human or animal learning.

Learning is a process consisting of the acquisition or transformation of knowledge, skills, behaviour, beliefs or orientations through experience. It is only a relatively stable change, resulting from a new single experience, or from the reiteration of the same experience. Learning may also involve the synthesis of different types of content.

Properties of learning processes. The acquisition of new modes of action or reaction, in the context of learning, may include both behaviour and emotional responses, not necessarily manifest, thus facilitating the individual's adaptation to environmental changes. By 'environment', we refer not only to the physical context in which the individual finds himself, but also to the set of variables, including relational ones, to which he is exposed. Learning is inherently an individual process, but it can also involve a social dimension based on sharing and cooperation, called collaborative learning. It is important to note that what is learned is not necessarily morally correct; the individual can assimilate both positive and negative habits. Moreover, learning processes are not always conscious: they may also occur unintentionally, for instance when acquiring skills while performing activities with other purposes in mind. The effects of learning processes are not always manifested through explicit actions; for example, acquired information, knowledge, skills and beliefs may remain part of an individual's personal baggage without necessarily translating into manifest behaviour. An individual may decide not to openly express what he or she has learnt. Interaction with

[187] Zhou, Z. H. (2021). Machine learning. Springer Nature.

[188] Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University press.

animals can be a vehicle for learning scientific content, such as the characteristics of the animal itself, but it can also foster emotional learning, e.g. compassion and respect.

Learning processes. The experiences we go through have a significant impact on neuronal connections, i.e. the messages exchanged between the cells that make up brain tissue, influencing both their functionality, through changes in the amount of neurotransmitters released, and their structure, through the expansion or reduction of connections. This phenomenon is known as neuronal plasticity and is particularly evident during childhood. Neuronal plasticity manifests itself in two main contexts:

1. During normal brain development, when the brain begins to process sensory information until adulthood, known as developmental plasticity and plasticity of learning and memory.
2. As a compensatory mechanism in response to loss of function and/or to maximise the potential of residual function after brain damage.

The environment plays a significant role in modulating neuronal plasticity: environments rich in stimuli appropriate to the child's developmental needs favour this phenomenon.

Examples from the natural learning of animals can be considered to better understand the cognitive processes of learning. Some of the most fundamental issues in machine learning already arise in that context, which is already known from a scientific point of view[189] . We refer, in particular, to the so-called 'Theory of Conditioning'. The phenomenon of conditioning is characterised by the process by which an association is established between an unconditioned, innate stimulus and a conditioned, artificial stimulus within an organism. In this context, the unconditioned stimulus spontaneously elicits a response, the conditioned stimulus, by its arbitrariness, exploits its proximity to evoke a similar response. The empirical emergence of conditioning is commonly attributed to the pioneering work of Russian physiologist Ivan Pavlov. Other scholars of the theory were John Watson and Burrhus Skinner.

Bait Shyness - Rats Learning to Avoid Poisonous Baits: When rats encounter food with a novel appearance or smell, they will first eat very small amounts of it and subsequent feeding will depend on the taste of the food and its physiological effect. If the food has a negative effect, the new food will often be associated with disease and, consequently, the rats will not eat it. Clearly, a learning mechanism is at play here: the animal uses past experience with a particular food to gain experience in detecting the safety of this food. If the past experience with the food has been negatively labelled, the animal predicts that it will also have a negative effect

[189] The examples are taken from the analysis by Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.

in the future. Based on this example of successful learning, one can consider a typical *machine learning* task. Suppose you want to programme a machine that learns to filter spam e-mails. A naive solution would apparently be similar to the way rats learn to avoid poisonous baits. The machine would simply store all previous e-mails that had been labelled as spam e-mails by the human user. When a new e-mail arrives, the machine will search for it in the set of previous spam e-mails. If matches one of them, it will be deleted. Otherwise, it will be moved to the user's inbox.

Pigeon Superstition: In an experiment conducted by the psychologist B. F. Skinner, he placed a group of starving pigeons in a cage. An automatic mechanism was attached to the cage that distributed food to the pigeons at regular intervals without any reference to the birds' behaviour. The hungry pigeons walked around the cage and, when the food was first delivered, found each pigeon engaged in some activity (pecking, turning its head, etc.). The arrival of the food reinforced each bird's specific action and, as a result, each bird tended to spend a little more time performing that same action. This, in turn, increased the chance that the next random delivery of food would find each bird engaged in that activity again. The result is a chain of events that reinforces the pigeons' association between the food delivery and the random actions they were performing at the time of the delivery. They then continue to diligently perform these same actions.

"What distinguishes the learning mechanisms that result in superstition from useful learning?" This question is crucial, for example, for the development of 'automated' learners. While human learners can rely on common sense to filter out random and meaningless learning conclusions, once we export the task of learning to a machine, we must provide crisp, well-defined principles that protect the programme from reaching meaningless or useless conclusions. The development of such principles is a central goal of machine learning theory. To understand the element that made rat learning more effective than pigeon learning, one can look more closely at the first phenomenon described in the examples.

Bait Shyness *revisited - rats fail to acquire conditioning between food and electric shock or between sound and nausea*: The mechanism of 'bait shyness' in rats turns out to be more complex than one might expect. In experiments conducted by Garcia (Garcia & Koelling 1996), it was shown that if the unpleasant stimulus that follows food consumption is replaced by, for example, an electric shock (instead of nausea), then no conditioning occurs. Even after repeated trials in which the consumption of some food was followed by the administration of unpleasant electric shocks, the rats did not tend to avoid that food. A similar failure of conditioning occurs when the feature of the food that implies nausea (such as taste or smell) is replaced

by a vocal signal. Rats seem to have some 'built-in' prior knowledge that tells them that while the temporal correlation between food and nausea may be causal, there is unlikely to be a causal relationship between food consumption and electric shocks or between sounds and nausea.

It can be inferred that a distinguishing feature between the learning of bait shyness and pigeon superstition is the incorporation of prior knowledge that influences the learning mechanism. This is also called inductive bias. The pigeons in the experiment are willing to adopt any explanation for the presence of the food. However, the rats 'know' that food cannot cause an electric shock and that the simultaneous presence of noise with some food cannot influence the nutritional value of that food. The rats' learning process is biased towards identifying certain types of patterns while ignoring other temporal correlations between events.

It is understood that the incorporation of prior knowledge, which influences the learning process, is inevitable for the success of learning algorithms[190] . The development of tools to express domain expertise, translating it into a learning *bias* and quantifying the effect of this *bias* on learning success is a central theme of machine learning theory. In other words, the stronger the prior knowledge (or prior assumptions) with which the learning process begins, the easier it will be to learn from further examples. However, the stronger these assumptions are, the less flexible learning is: it is bound, a priori, by the commitment to these assumptions. Depending on the learning approach, the type of data they input and output and the type of problem they solve, there are a few main categories of supervised, unsupervised and reinforcement learning machine learning algorithms. There are some hybrid approaches and other common methods that offer a natural extrapolation of the problematic forms of machine learning[191] . Depending on the desired outcome of the algorithm, machine learning algorithms are organised into the following groups .[192]

Supervised learning. The definition of supervised learning is illustrated as a block diagram in Fig. 1.1. One can consider the two types of output value in supervised learning: the target output initially assigned to each training example and the calculated output that is computed by the learning algorithm. Supervised learning is the learning paradigm in which the parameters are optimised to minimise the difference between the target output and the calculated output. The KNN (K Nearest Neighbour), the Naive Bayes and the decision tree

[190] This is the case with the so-called 'free lunch theorem'.

[191] Sah, S. (2020). Machine learning: a review of learning types.

[192] For learning types, see Jo, T. (2021). Machine learning foundations. Machine Learning Foundations. Springer Nature Switzerland AG. https://doi. org/10.1007/978-3-030-65900-4.

belong to supervised learning; moreover, areas of application of this type of learning are classification and regression. There are several training examples that learn the supervised learning algorithms. Each training example is associated with its own category or continuous value. The output initially labelled for each training example is called the target output, while the output calculated by the learning algorithm is called the calculated output. The objective of supervised learning is to minimise misclassification or error between the target output and the calculated output. Unlabelled or *labelled* examples are added to labelled examples prepared for supervised learning, in the type of learning called semi-supervised learning. The learning process of labelled training examples in supervised learning algorithms should also be mentioned. Their parameters are randomly initialised and the labelled training examples are prepared. Their output values are calculated and the parameters are updated to minimise the error between the target outputs and the calculated outputs. The process is repeated to minimise the error or achieve parameter convergence. The learning process in supervised learning consists of optimising the parameters to minimise error. The main feature of supervised learning is the predetermination of classes. These classes are created in a kind of finite, human-defined set, which in practice means that a certain segment of data will be labelled with these classifications. The task of the machine learning algorithm is to find patterns and build mathematical models. These models are then evaluated for predictive ability against the variance measures in the data itself. In the learning process, the parameters are optimised to minimise the misclassification rate or error. Classification and regression are the tasks to which supervised learning algorithms are applied. The case where unlabelled training examples are used together with labelled examples for the learning process is called semi-supervised learning. It is also useful to distinguish between two main models of supervised learning: classification models (classifiers) and regression models. The latter map the input space into a domain of real values. Classifiers map the input space into predefined classes. There are many alternatives for representing classifiers, e.g. support vector machines, decision trees, probabilistic summaries, algebraic functions. Together with regression and probability estimation, classification is one of the most studied models, perhaps the one with the greatest practical relevance. The potential benefits of advances in classification are immense as the technique has a great impact on other areas, both in data mining and its applications.

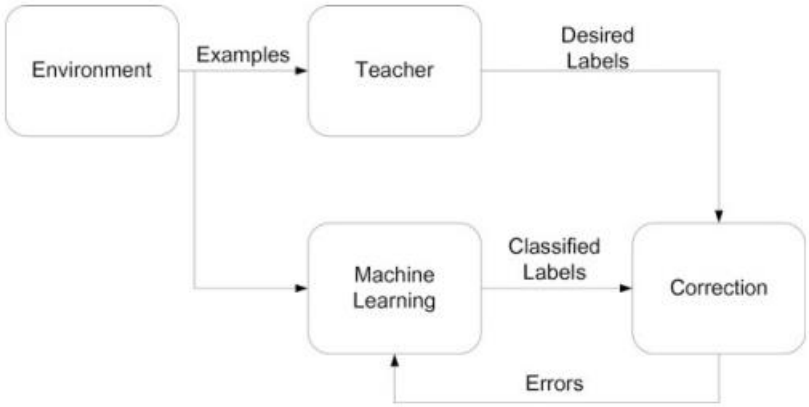

Fig. 1

Unsupervised learning. Unsupervised learning as a further learning paradigm is illustrated as a block diagram in Fig. 1.2. The training examples provided for the learning paradigm are not labelled and the clustering prototypes are randomly initialised. Unsupervised learning is defined as the process of optimising the cluster prototypes, depending on the similarities between the training examples. Clustering of data elements is the task to which unsupervised learning algorithms are applied. We cite the training examples prepared for unsupervised learning. In unsupervised learning, all training examples are unlabelled, and the prototypes, each of which characterises its own cluster, are randomly initialised. The unlabelled training examples are used to optimise the cluster prototypes according to their similarities. The case of combining unlabelled examples with some labelled examples for clustering data elements is called clustering with constraints. It also appears necessary to note that the learning process in the unsupervised learning algorithm takes the form of a framework rather than a specific algorithm. The number of clusters is decided and cluster prototypes are initialised. Each data element in the cluster is organised into the cluster whose prototype is most similar and are updated into more similar elements based on the organised elements won. The arrangement and updating of cluster prototypes are iterated until they converge. Supervised learning serves to minimise errors or misclassifications between computed and target outputs, while unsupervised learning serves to maximise the similarities between cluster prototypes and data elements. Whereas in supervised learning, training examples are labelled with their own output values, in unsupervised learning they are not labelled. Supervised learning depends on the target labels of the training examples, whereas unsupervised learning depends on their similarities to the cluster prototypes. Supervised learning algorithms are applied to classification and regression, while unsupervised learning algorithms are applied to clustering.

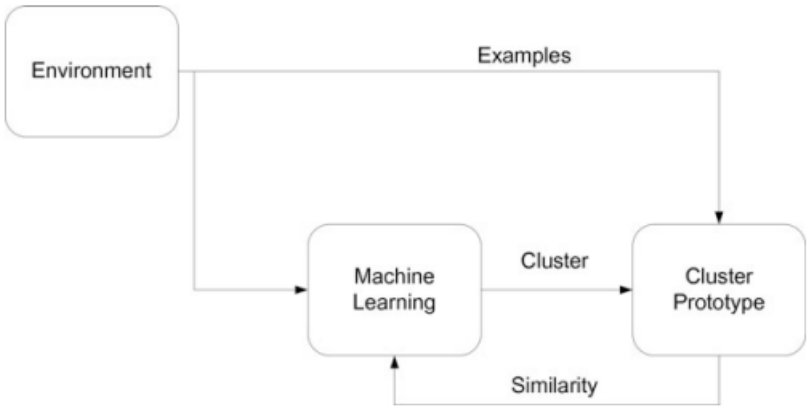

Fig. 2

Semi-supervised learning Semi-supervised learning is illustrated as a block diagram in Fig. 1.3. It is intended to use unlabelled examples that are very cheap to obtain as well as labelled examples to train the learning algorithms. Machine learning learns the labelled examples by minimising the error between the target and computed labels and the unlabelled examples, depending on their similarities, such as the mixture of supervised and unsupervised learning. By adding the unlabelled examples, semi-supervised learning is applied to classification tasks and regression tasks. Semi-supervised learning can be described as the combination of both types of learning. The fact that labelled examples are expensive to obtain but unlabelled examples are cheap to produce is the motivation for proposing semi-supervised learning. It is intended to add the unlabelled examples to the labelled ones in order to strengthen classification and regression performance. In semi-supervised learning, the training set is divided into two sets: the set of labelled examples and the set of unlabelled examples. The originally unlabelled examples are artificially labelled based on their similarities to the originally labelled examples or by the learning classifiers. Semi-supervised learning is intended for tasks to which supervised learning algorithms, such as classification and regression, are applied. The originally labelled training examples are prepared and the unlabelled ones are added. The parameters of the machine learning algorithms are updated by minimising the error between computed labels and target labels in labelled examples and minimising the distance between cluster prototypes and input vectors in unlabelled examples. Semi-supervised learning is implemented by combining a supervised and an unsupervised learning algorithm; unlabelled training examples are clustered by the unsupervised learning algorithm, based on labelled examples, and both originally labelled and subsequently labelled examples are used to train the supervised learning algorithm. Semi-supervised learning is seen as a composite or combination of supervised and unsupervised learning. Semi-supervised learning is implemented by combining the unsupervised and supervised learning algorithm.

The labelled and unlabelled examples are initially provided as training examples, and the unlabelled examples are grouped by the unsupervised learning algorithm, referring to the labelled ones. The supervised learning algorithm is trained with both types of examples. One can, at this point, reflect on the case of clustering data elements using both types of examples. In unsupervised learning, examples are assumed to be unlabelled. The clustering of data elements in which external parameters are optimised by referring to labelled examples and unlabelled examples are clustered is called constraint clustering. Although constraint clustering is included as part of the semi-supervised learning process, it should be distinguished from semi-supervised learning with respect to their objectives. The fact that labelled examples are provided as constraints in the clustering of data elements is the reason why it is called task constraint clustering.

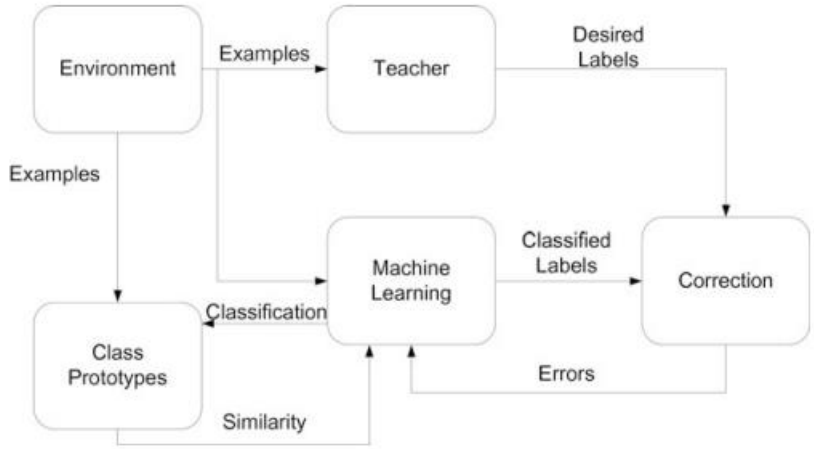

Fig. 3

Reinforcement learning. Reinforcement learning - schematised in Fig. 1.4 - is defined as the type of learning in which parameters are updated to maximise the reward and minimise the penalty. In reinforcement learning, the input is received from the external environment and the output is generated as an action. The reward or penalty is given by the environment as critical, and the parameters are updated to obtain the reward and avoid the penalty as far as possible. One must consider the external environment to study reinforcement learning. This type of learning is seen as the interaction between the environment and the learning agent. The reward or penalty is decided by the environment after the action is taken by the learning agent. The environment strongly influences the learning agent's decision which action is most desirable for perception, it plays the role of a critic who judges the learning agent's action. In other words, the learning agent accepts the perception of the external environment as input and generates the action randomly or from its table as output. The action is given the reward or penalty. If the reward is given, the current output is set as the desired output and if not, the output is updated. The implementation of the autonomous moving agent is a typical area

where the learning type is used. The pending type is applied to implement a game player that is the opposite of the users. We might consider applying reinforcement learning to stochastic classification in which the target labels of the training examples are not fixed. There is a possibility to apply reinforcement learning to implement the autonomous commercial agent in the financial area. Let us make some remarks on reinforcement learning. In supervised learning and unsupervised learning, it is assumed that the training examples are fixed, but in this type of learning, the training examples are variable. As time passes, the reward and penalty even for the same input are variable.

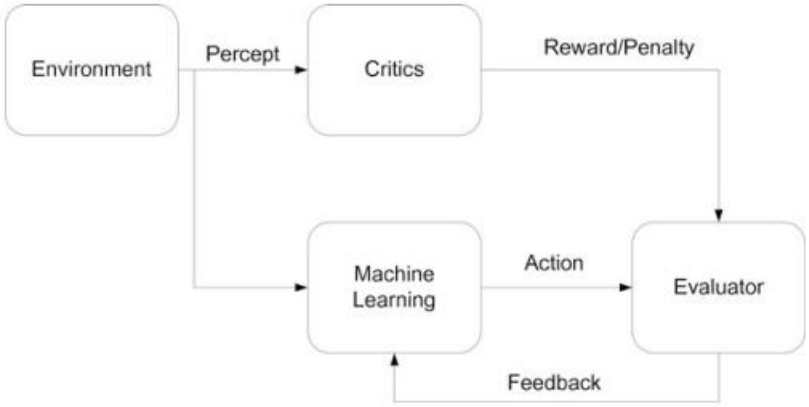

Fig. 4

Deep learning[193] uses multilayer neural networks trained with large data sets to solve complex information processing tasks and has emerged as the most successful paradigm in the field of *machine learning*. The current phase of neural network development began during the second decade of the 21st century. A series of developments allowed neural networks with many layers of weights to be trained effectively, thus removing previous limitations on the capabilities of these techniques. Networks with many layers of weights are called deep neural networks and constitute the sub-field of machine learning that focuses on such networks, *deep learning*. Fundamental to the development of deep learning has been a significant increase in the scale of neural networks, measured in terms of the number of parameters. Although networks with a few hundred or a few thousand parameters were common, in the 1980s, this number steadily increased to millions, and then billions, while today's state-of-the-art models can have around a trillion (1012) parameters. Networks with many parameters require proportionately large data sets so that good values can be produced for these parameters. The combination of massive models and massive data sets in turn requires large-scale calculations during model training. Specialised processors called graphics processing units or

[193] Bishop, C. M., & Bishop, H. (2023). Deep learning: Foundations and concepts. Springer Nature.

GPUs, which had been developed for very fast rendering of graphical data for applications such as video games, turned out to be suitable for training neural networks because the functions computed by the units in one layer of a network can be evaluated in parallel and this fits well with the massive parallelism of GPUs (Krizhevsky, Sutskever and Hinton, 2012). Today, training for larger models is performed on large arrays of thousands of GPUs connected by specialised high-speed interconnects. replaced by simply increasing the amount of training data, along with a proportionate scaling of model size and associated computing power used for training (Sutton, 2019). Not only can large models have superior performance on a specific task, but they may be able to solve a wider range of different problems with the same trained neural network. Large language models are a notable example as a single network not only has an extraordinary breadth of capability, but is also able to outperform specialised networks designed to solve specific problems. Depth plays an important role in enabling neural networks to achieve high performance. One way of looking at the role of hidden layers in a deep neural network is through representation learning (Bengio, Courville and Vincent, 2012) in which the network learns to transform input data into a new semantically meaningful representation thus creating a much easier problem for the final layer(s) to solve. Such internal representations can be repurposed to enable the solution of related problems through transfer learning. Interestingly, the neural networks used to process images can learn internal representations remarkably similar to those observed in the mammalian visual cortex. Large neural networks that can be adapted or tuned for a range of downstream tasks are called basic models and can take advantage of large and heterogeneous datasets to create models with broad applicability (Bommasani et al., 2021).

Besides scalability, there have been other developments that have contributed to the success of deep learning. For example, in simple neural networks, training signals become weaker as they are back-propagated through successive layers of a deep network. One technique to address this problem is the introduction of residual connections (He et al., 2015a) that facilitate the training of networks with hundreds of layers. Another key development has been the introduction of automatic differentiation methods in which the code that performs backward propagation to evaluate the gradients of the error function is automatically generated from the code used to specify forward propagation. This allows researchers to quickly experiment with different architectures for a neural network and to combine different architectural elements in multiple ways very easily since only relatively simple forward propagation functions need to be coded explicitly. Furthermore, much of the machine learning research has been conducted via open source, allowing researchers to build on the

work of others, thus further accelerating the rate of progress in the field.

As mentioned, one of the most important advances in the field of machine learning in recent years has been the development of powerful models for processing natural language and other forms of sequential data such as source code. A *Large Language Model*, or LLM, uses deep learning to construct rich internal representations that capture the semantic properties of language. An important class of large language models, called autoregressive language models, can generate language as output and, therefore, are a form of generative artificial intelligence. Such models take a sequence of words as input and as output generate a single word representing the next word in the sequence. The augmented sequence, with the new word added at the end, can then be fed back into the model to generate the next word and this process can be repeated to generate a long sequence of words. Such models can also produce a specific word that signals the end of text generation, allowing them to produce text of finite length and then stop. At that point, a user can add his or her own set of words to the sequence before feeding the complete sequence back into the template to trigger further word generation. In this way, it is possible for a human being to converse with the neural network. Such models can be trained on large text datasets by extracting training pairs, each consisting of a randomly selected sequence of words as input with the next known word as the target output. This is an example of self-supervised learning in which a function is learned from inputs to outputs but in which labelled outputs are obtained automatically from the input training data without the need for separate human-derived labels. Since large volumes of text are available from multiple sources, this approach allows for scaling very large training sets and associated very large neural networks. Large language models can display extraordinary capabilities that have been described as the first indications of emerging general artificial intelligence (Bubeck et al., 2023).

1.1 *Foundation Models* and *Large Language Models*.

Artificial Intelligence in its evolutionary path has recently seen a real revolution that has led to a radical paradigm shift, represented by the so-called *Foundation Models*. A *Foundation Model* is any model trained on big data (usually using large-scale self-supervision) that can be adapted to a wide range of downstream tasks; current examples include BERT [Devlin et al. 2019], GPT-3 [Brown et al.2020] and CLIP [Radford et al. 2021]. From a technological point of view, the basic models are not new: they are based on deep neural networks and self-paced learning, both of which have existed for decades. However, the breadth and scope of

foundation models in recent years have broadened our imagination of what is possible; for example, GPT-3 has 175 billion parameters and can be adapted via natural language instructions to do an acceptable job on a wide range of tasks despite not being explicitly trained to perform many of these tasks [Brown et al. 2020]. At the same time, existing *Foundation Models* have the potential to exacerbate damage and their characteristics are generally poorly understood. Given their imminent spread, they have become a topic of intense scrutiny [Bender et al. 2021]. The significance of foundation models can be summarised in two words: emergence and homogenisation. Emergence means that the behaviour of a system is implicitly induced rather than explicitly constructed. Homogenisation means the consolidation of methodologies for building machine learning systems across a wide range of applications; it provides strong leverage towards many tasks but also creates single points of failure.[194]

Technically, *language modelling* (LM) is one of the main approaches to advance the linguistic intelligence of machines and thus a *Foundation Model* trained on textual data. In general, LM aims at modelling the generative probability of word sequences in order to predict the probability of future (or missing) tokens.[195] The most recent development of this approach

[194] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S.. & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

[195] On this point, see Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223. In particular, with reference to language modelling, the authors highlight a historical differentiation from its beginnings up to Large Language Models. First of all, Statistical Language Models are developed on the basis of statistical learning methods popular in the 1990s. The basic idea is to build a word prediction model based on the Markov hypothesis, i.e. to predict the next word based on the most recent context. SLMs with a fixed context length n are also called n-gram language models, e.g. bigram and trigram language models. SLMs have been widely applied to improve the performance of information retrieval (IR) and natural language processing (NLP) tasks. However, an issue related to dimensionality is often encountered: it is difficult to accurately estimate high-order linguistic models since an exponential number of transition probabilities must be estimated. Therefore, specially designed levelling strategies such as backoff estimation and Good-Turing estimation were introduced to alleviate the problem of data sparsity. Next, Neural Language Models (NLM) were introduced. They characterise the probability of word sequences by neural networks, e.g. multilayer perceptron (MLP) and recurrent neural networks (RNN). As a notable contribution, the concept of distributed word representations was introduced and word prediction functions conditioned on aggregate context features (i.e. distributed word vectors) were constructed. Extending the idea of learning effective features for text data, a general neural network approach was developed to build a unified end-to-end solution for various NLP tasks. Furthermore, word2vec was proposed to build a simplified surface neural network for learning distributed word representations, which proved to be very effective in a variety of NLP tasks. These studies initiated the use of language models for representational learning (beyond word sequence modelling), having an important impact in the field of NLP. With reference to the latter, Pre-trained Language Models, as a first attempt, ELMo was proposed to acquire context-sensitive word representations by pre-training a bi-directional LSTM (biLSTM) network (instead of learning fixed word representations) and then refining the biLSTM network according to specific downstream tasks. Furthermore, based on the highly parallelisable Transformer architecture with self-attention mechanisms, BERT was proposed using bi-directional pre-training language models with specially designed pre-training tasks on large-scale label-free corpora. These pre-trained, context-sensitive word representations are highly effective as general-purpose semantic features, which have largely raised the performance level of NLP tasks. This study has inspired a large number of follow-up works, which call the learning paradigm 'pre-training and refinement'. Following this paradigm, a large number of PLM studies have been developed, introducing different

can be found in the so-called. *Large Language Models* (LLM). These take the form of artificial neural networks pre-trained on huge datasets which, instead of being dedicated to a specific task, can perform a wide range of different functions at a level that is all the more accurate the more resources (e.g. datasets and computing power) are made available to them[196] . The concept of *language models* is not new but has evolved with the progress of artificial intelligence over the decades. Early language models mainly aimed at modelling and generating text data, whereas more recent language models (e.g. GPT-4) focus on solving complex tasks. However, the shift from language *modelling* to *task solving* is crucial in scientific thinking, which is the key to understanding the development of language models in the history of research. From the point of view of task solving, the four generations of language models showed different levels of model capability. Initially, statistical language models mainly helped in some specific tasks, where predicted or estimated probabilities can improve the performance of task-specific approaches. Subsequently, neural language models focused on learning task-independent representations, with the aim of reducing efforts for human feature engineering. Furthermore, pre-trained language models learned context-sensitive representations that can be optimised according to downstream tasks. For the latest generation of language models, LLMs are improved by exploring the effect of scalability on the model's capability, which can be considered as a solver of generic tasks. In other words, in the process of evolution, the scope of tasks that can be solved by language models has been significantly extended, and the performance of tasks achieved by language models has been significantly improved[197] . While the previous models, and especially the *Pre-trained Language Models*, have been explored and treated in detail, in contrast, the *Large Language Models* are only now beginning to be discussed and analysed.

In order to better define and understand the scope of *Large Language Models,* some authors[198]

architectures (e.g. GPT-2 and BART) or improved pre-training strategies. In this paradigm, it is often necessary to fine-tune the PLM to suit different downstream activities. Subsequently, researchers find that scaling of PLMs (e.g. scaling of model size or data) often leads to improved model capability in downstream activities (i.e., following the scaling law). Numerous studies have explored the performance limit by training an increasingly larger PLM (e.g. the GPT-3 with parameter 175B and the PaLM with parameter 540B). Although scaling is mainly done according to model size (with similar architectures and pre-training activities), these larger PLMs show different behaviour compared to smaller PLMs (e.g., BERT with parameter 330M and GPT-2 with parameter 1.5B) and exhibit surprising abilities (called emergent abilities) in solving a variety of complex tasks. For example, GPT-3 can solve tasks with few operations through context learning, whereas GPT-2 cannot do well. Therefore, the research community coined the term 'Large Language Models (LLM)' for these large PLMs, which attract increasing research attention.

[196] h ttps://www.agendadigitale.eu/industry-4-0/ia-limpatto-dei-large-language-model-sulle-aziende-usi-e-problemi/

[197] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

[198] See also Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

, who were the first to attempt to establish criteria to differentiate PLMs from LLMs, suggest looking at some of the latter's distinctive features. In particular, LLMs display certain emergent capabilities that might not be observed in previous smaller PLMs[199] . Second, LLMs would revolutionise the way humans develop and use artificial intelligence algorithms[200] . Thirdly, the development of LLMs no longer draws a clear distinction between research and engineering. Indeed, the training of LLMs requires extensive practical experience in large-scale data processing and parallel distributed training[201] . The implementation of LLMs would potentially lead to a thriving ecosystem of real-world applications based on them. Despite the progress and impact, the principles behind LLMs are not yet well explored. Firstly, it is not well understood why emerging skills occur in LLMs, rather than in smaller PLMs. From a more general point of view, there is a lack of in-depth and detailed investigation of the key factors that contribute to the superior capabilities of LLMs. Indeed, due to the huge demand on computing resources, it is very expensive to conduct in-depth studies to verify the effect of various strategies for training LLMs. Indeed, LLMs are mainly trained in industrial settings where many important details of training (e.g. data collection and cleaning) are not revealed. Third, it is difficult to align LLMs with human values or preferences. Despite their capabilities, LLMs are also likely to produce toxic, fictitious or harmful content. Therefore, effective and efficient control approaches are required to eliminate the potential risk arising from the use of LLMs. Faced with both opportunities and challenges, more attention is needed in the research and development of LLMs, which could be understood more deeply by systematically analysing four aspects: pre-training, adaptation, utilisation and evaluation of capabilities.

Generally, LLMs refer to Transformer language models containing hundreds of billions (or more) of parameters that are trained on a massive amount of data. LLMs show a strong ability to understand natural language and solve complex tasks (via text generation). This is also made possible by the use of deep neural networks. Below are the four aspects considered fundamental in the architecture of LLMs.

Pre-training. Pre-training establishes the foundation of the LLMs' capabilities. Through *pre-training* on large-scale corpora, LLMs can acquire essential language understanding and

[199] These capabilities are crucial for the execution of language models in complex tasks, making artificial intelligence algorithms extraordinarily powerful and effective.

[200] Unlike small PLMs, the main approach for accessing LLMs is via the request interface (e.g. API GPT-4). Humans must understand how LLMs work and format their activities so that LLMs can follow them.

[201] To develop capable LLMs, researchers have to solve complicated engineering problems, working with engineers or being engineers. Nowadays, LLMs are having a significant impact on the AI community and the advent of ChatGPT and GPT-4 leads to rethinking the possibilities of artificial general intelligence.

generation skills. In this process, the scope and quality of the pre-training corpus are crucial for LLMs to acquire powerful capabilities. Furthermore, to effectively pre-train LLMs, model architectures, acceleration methods and optimisation techniques must be well designed.

Adaptation. After *pre-training*, LLMs can acquire general capabilities to solve various tasks. However, an increasing number of studies have shown that the capabilities of LLMs can be further adapted according to specific objectives. There are two main approaches to adapt pre-trained LLMs, namely instructional optimisation and alignment optimisation. The first approach mainly aims at improving (or unlocking) the capabilities of LLMs, while the second approach aims at aligning LLMs' behaviours with human values or preferences. Furthermore, it is important to pay attention to efficient tuning and quantisation for model adaptation in resource-limited contexts. Generally, an *instruction-formatted* instance consists of a task description (called an instruction), an optional input, the corresponding output and a small number of (optional) demonstrations. Although a large number of instances are of this type, they mainly come from public NLP datasets, lacking instruction diversity or not corresponding to real human needs. To overcome this problem, InstructGPT for instance proposes to take queries that real users have sent to the OpenAI API as task descriptions. Furthermore, to enrich the diversity of tasks, human taggers are also asked to compose instructions for real-life tasks, including open generation, answering open-ended questions, brainstorming and chatting. They then let another group of labellers respond directly to these instructions as output. Finally, they couple an instruction (i.e. the collected user query) and the expected output (i.e. the response written by a human) as a training instance. Note that InstructGPT also employs these real-world activities formatted in natural language for alignment tuning. Furthermore, GPT-4 designed potentially high-risk instructions and guided the model to reject these instructions through supervised tuning for safety issues. Considering the absence of high-quality public chat data, several studies also collected user chat requests as input data and then used ChatGPT or GPT-4 to generate responses as output data. A notable example of such a dataset is the ShareGPT chat data. In addition, Dolly and OpenAssistant further released their conversation data, which was carefully labelled by human annotators to achieve a high level of quality.

Usage. After pre-training or fine-tuning the adaptation, an important approach to using LLMs is to design appropriate suggestion strategies to solve various tasks. In the existing literature, task-specific hints can be learnt effectively through manual creation and automatic optimisation. One representative suggestion method is context learning, which formulates task descriptions and/or demonstrations in the form of natural language text. Furthermore,

chain thinking-based hints can be used to enhance learning in context by involving a series of intermediate reasoning steps in the hints. Furthermore, planning is proposed for solving complex tasks, which first breaks them down into smaller subtasks and then generates an action plan to solve these subtasks one by one.

Capacity assessment. In order to examine the effectiveness and superiority of LLM, a wave of tasks and benchmarks have been proposed to conduct empirical capability assessments and analyses. Different types of assessment of the basic skills of LLMs for language generation and comprehension, thus different are the resulting assessments of advanced skills with more complicated settings or goals.

2. The European Union Regulation on Artificial Intelligence: a new approach to digital?

European standardisation in the field of digital and automation, especially in the aftermath of the increasingly massive and thus necessary use of Artificial Intelligence, took its start in 2020 from the so-called Data Strategy 2030[202] . Launched by the European Commission, the document set itself the goal of devising a uniform regulatory system applicable at European level, aimed at regulating the data economy and preventing the risks and abuses arising from the dominant position of large online platforms. The regulatory choices expressed by the European legislator within the Strategy for the Data Economy focus, inter alia, on crucial issues such as the availability, interoperability and sharing of data, the definition of clearer modalities concerning the use of data, as well as on the governance of the related processes and the creation of common spaces at European level. In this sense, the European Strategy should be read as an instrument through which the European Union intends to undertake concrete legislative action aimed at creating a single European data market that respects the fundamental principles on which the Union is founded. The initiative is part of a context of strengthening internal cohesion and promoting balanced regulation that, while safeguarding competitiveness, pays attention to the protection of users' rights and the mitigation of competitive distortions generated by dominant platforms[203] . Given this evolution in the regulatory sense, new demands for protection and balance continue to emerge that look, on

[202]European, C. (2020). A European data strategy. COM (2020), 66. See also Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19 , p. 120.

[203] Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19. P. 120-121; see also Fabiano, L. (2024). Algorithmic Recommendation and Artificial Intelligence in the US and European experience between strategic needs and fundamental rights protection. IRCOCERVO, 23(1), 440-456.

the one hand, to the clear advantages of the use of automation, as well as the data market; on the other hand, to the protection of fundamental principles and rights concerning the human person. For these reasons, the European institutions have enacted a package of rules - all in the form of a regulation, thus aiming at full harmonisation - governing the data market and automation, considering different aspects. As mentioned (see *above*), the European legislator was the first on a global scale to regulate Artificial Intelligence with EU Regulation 2024/1689, following an initial proposal dated April 2021. The intention of the Regulation - in the wake of the Artificial Intelligence White Paper[204] - is twofold: to foster the adoption of artificial intelligence technology and to address the risks associated with it. The main intention is, therefore, to create an ecosystem based on trust, through the establishment of a legal framework to ensure the development of trustworthy artificial intelligence. Generally speaking, the Regulation stipulates that the European regulation of artificial intelligence should be characterised by a balanced and proportionate approach, avoiding excessive restrictions on the development of new technologies, but at the same time highlighting the potential risks associated with certain uses of artificial intelligence systems[205] . From a legal point of view, it is necessary to ask whether the structure of the regulation is appropriate to meet the requirements that permeate it. Indeed, the question must be asked whether the approach adopted by the legislator, regulating AI as a whole, is more effective than an approach that would otherwise have regulated artificial intelligence applications in various fields. The regulatory approach followed by the AI Act was a horizontal one, based on a broad definition of artificial intelligence. In particular, Article 3 of the Regulation[206] includes in the concept of artificial intelligence those systems which, on the basis of explicit or implicit objectives, deduce from the input received how to generate output such as predictions, content, recommendations or decisions capable of influencing physical or virtual environments, excluding, instead, less complex software and AI systems used exclusively for military, defence, research and innovation, or non-professional purposes. The definition of an automated system, consistent with the one proposed in the OECD[207] , is deliberately

[204]White Paper on Artificial Intelligence - A European Approach to Excellence and Trust [COM(2020) 65 final].
[205] Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19. P. 120-121; see also Fabiano, L. (2024). Algorithmic Recommendation and Artificial Intelligence in the US and European experience between strategic needs and fundamental rights protection. IRCOCERVO, 23(1), 440-456.p.121.
[206] Art. 3 AI ACT .
[207] OECD (2019), Artificial Intelligence in Society, OECD Publishing, 11 June 2019: 'system that can, for a given set of humandefined objectives, make predictions, recommendations or decisions influencing real or virtual environments. It uses machine and/or humanbased inputs to perceive real and/or virtual environments; abstract such perceptions into models (in an automated manner e.g. with ML or manually); and use model

broad and generic, with the intention of being as neutral as possible and reducing the risk of rapid obsolescence of regulation. The choice of a horizontal regulatory approach also has relevant geopolitical implications in the international context. As stated in the explanatory memorandum to the regulation, the EU's main objective is to 'protect the Union's digital sovereignty and use its regulatory tools and powers to shape rules and standards with global reach'. Therefore, the EU's strategy aims to make the European model a global reference point, to be implemented worldwide, as has already happened in the past with Regulation (EU) No 2016/679 (GDPR). In fact, on an international level, characterised by the absence of an organic discipline, the European Regulation stands as the only legislation that is not fragmentary and incomplete, as is the case with the regulations of China and the United States, the main competitors in the artificial intelligence sector. Moving the perspective of analysis towards the limits and criticalities of the Regulation, one can point out how a technical-legal limit of the approach adopted by the European legislator concerns the fact that the rules of the Regulation do not solve specific problems or fill precise gaps in the legal system, but are applied in various sectors of civil society[208] . As regards, on the other hand, the subjective scope of the AI Act[209] , it applies to providers who place on the market or put into service artificial intelligence (AI) systems or AI models for general purposes in the European Union, irrespective of their location, be it in the Union or in a third country. It also extends to *deployers* of AI systems established or located in the Union, as well as suppliers and *deployers* established or located in third countries, if the output produced by the AI system is used in the Union. The scope also includes importers and distributors of IA systems, manufacturers of products that place an IA system on the market together with their product, under their own name or trademark, and authorised representatives of suppliers not established in the Union. Finally, it applies to natural and legal persons who are located in the Union and are affected by the use of AI systems. The regulation also takes particular account of AI systems that are qualified as "high risk"[210] and other areas of law: in this case,

inference to formulate options for information or action. AI systems are designed to operate with varying levels of autonomy'.

[208] On this point, see more extensively Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19. P. 120-121; see also Fabiano, L. (2024). Algorithmic Recommendation and Artificial Intelligence in the US and European experience between strategic needs and fundamental rights protection. IRCOCERVO, 23(1), 440-456, pp.121-122.

[209] See Art. 2 AI ACT

[210] The Art.6 para.1 AI ACT states: "1. Regardless of whether it is placed on the market or put into service independently of the products referred to in points (a) and (b), an AI system shall be considered high-risk if both of the following conditions are met

a)the IA system is intended to be used as a safety component of a product, or the IA system is itself a product, covered by Union harmonisation legislation listed in Annex I

the regulation will apply in part, i.e. it will not apply[211] . The Regulation, the text of which consists of 113 articles and 13 annexes (so-called annex technique with a periodic five-yearly review)[212] , is based on a *risk-based* architecture, whereby it is possible to distinguish between different risk categories.

Unacceptable Risk Systems. Under Section 5(1) AI Act, all practices such as the placing on the market, commissioning or use of an AI system using subliminal techniques that operate without the knowledge of a person, or manipulative or deceptive techniques that have the purpose or effect of substantially altering the behaviour of a person or group of persons, significantly impairing their ability to make informed decisions and causing them to make choices they would not otherwise have made, with harmful consequences for them or others, are prohibited. Also prohibited is the use of AI systems that exploit specific vulnerabilities of individuals or groups, such as those due to age, disability, or particular social or economic conditions, with the intent of altering behaviour in a way that may cause significant harm. In addition, the marketing or use of AI systems designed to evaluate or classify individuals or groups on the basis of social behaviour or personal characteristics is prohibited if the social score resulting from such classification would result in unjustified or disproportionate treatment of, or harm to, individuals in social contexts unrelated to those in which the data were originally collected. The use of AI systems for risk assessment of individuals, with the intention of predicting the likelihood of an individual committing crimes, based solely on profiling or the assessment of personal traits and characteristics is also prohibited, unless they are systems that support human assessment based on concrete and verifiable facts related to criminal behaviour. Furthermore, the use of AI systems that create or extend facial recognition databases by untargeted *scraping* of facial images from the Internet or CCTV footage is prohibited, as is the use of AI to infer emotions in work or educational contexts, unless such use is strictly for medical or security reasons. The use of biometric categorisation systems that classify individuals on the basis of biometric data to infer information relating to race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation is also prohibited, except in cases of legitimate labelling or filtering of biometric data for law enforcement purposes. Finally, the use of 'real-time' remote biometric

b)the product, the safety component of which in accordance with point (a) is the IA system, or the IA system itself being a product, is subject to a third-party conformity assessment for the purposes of placing that product on the market or putting it into service covered by Union harmonisation legislation listed in Annex I."

[211] See, again, Art. 2 AI ACT

[212] Christmas, G. (2024). The new European AI Act regulation. Internet law: Artificial intelligence, digital copyright and data protection, 2, 123-145.

identification systems in public spaces for law enforcement purposes is prohibited, except where such use is strictly necessary for the search of victims of human trafficking or the location of missing persons, for the prevention of an imminent threat to life or limb, or for the identification of a person suspected of serious crime, in the context of criminal investigations, prosecutions or executions of criminal offences punishable by deprivation of liberty for a term of not less than four years.

High-risk Artificial Intelligence Systems. These are those systems that Article 6(1) of Regulation[213] defines as such if they fulfil two concomitant conditions, regardless of whether they are placed on the market or put into service. The first condition is that the IA system is intended to be used as a safety component within a product, or that the IA system itself is a stand-alone product, in both cases subject to the regulations of the European Union harmonisation legislation set out in Annex I. The second condition stipulates that the product, which includes a safety component constituted by the IA system in question, or the system itself, is subject to a conformity assessment by third-party bodies, with a view to its being placed on the market or put into service, as provided for by the provisions of the Union harmonisation legislation listed in Annex I. High-risk systems are also defined as[214] those listed in the different sectors, in accordance with the relevant EU or national legislation permitting their use. In the area of biometrics, remote biometric identification systems are included, with the exclusion of AI systems intended exclusively for biometric verification to confirm the identity of a natural person. Also included are AI systems intended for biometric categorisation based on sensitive attributes or characteristics, as well as those used for emotion recognition. In the field of critical infrastructures, AI systems used as security components in the management and operation of critical digital infrastructures, road traffic, and the provision of essential services such as water, gas, heating and electricity are considered high-risk. With regard to education and vocational training, high-risk AI systems include those intended to determine the access, admission or assignment of individuals to educational or vocational training institutions at any level. Also included are systems used to assess learning outcomes, guiding the educational process, as well as those used to monitor and detect prohibited behaviour during testing in such contexts. In the area of employment, labour management and access to self-employment, AI systems used for recruitment and selection of personnel are considered high-risk, in particular for posting targeted job

[213] Art. 6(1) of EU Regulation 2024/1689

[214] Paragraph 2 of Art. 6 of the Regulation expressly refers to Annex III, which provides a detailed list of high-risk AI systems.

advertisements, analysing or filtering applications and assessing candidates. In addition, systems used to make decisions on the terms of employment relationships, the promotion or termination of contracts, the assignment of tasks, and to monitor and evaluate the performance and behaviour of workers fall into this category. With regard to access to essential private services and essential public benefits, AI systems intended for use by or on behalf of public authorities to assess the eligibility of individuals for essential public benefits, such as health care, and to determine the creditworthiness and assess the solvency of individuals are included, excluding systems used to detect financial fraud. Also included are AI systems used to determine risks and prices in relation to natural persons, in the insurance area, and those used to evaluate and classify emergency calls. In the area of law enforcement, this includes AI systems used by law enforcement authorities or EU institutions to determine the risk that a natural person may become a victim of crime. It also includes systems used to assess the reliability of evidence, to determine the risk of committing a crime or reoffending, as well as those used to perform profiling of natural persons during the investigation or prosecution of crimes. In the context of migration, asylum and border control management, AI systems used to assess risks associated with natural persons intending to enter the territory of a Member State, to assist in the examination of applications for asylum, visas or residence permits, as well as to detect, recognise or identify natural persons during the migration process, with the exception of the verification of travel documents, are considered high-risk. Finally, in the area of administration of justice and democratic processes, AI systems used by judicial authorities to assist in fact-finding, law interpretation and enforcement, as well as those used in alternative dispute resolution, are included. Also included are AI systems used to influence the outcome of an election or referendum, or the voting behaviour of voters, with the exclusion of those operating at an administrative and logistical level, such as tools for organising and optimising political campaigns. This list does not, however, represent a *numerus clausus* of options. In fact, Article 7 of the Regulation specifies that the European Commission is empowered to adopt delegated acts pursuant to Article 97[215] in order to amend Annex III, with the inclusion or modification of the use cases of high-risk artificial intelligence (AI) systems, if two conditions are met. Firstly, AI systems must be intended for use in one of the areas listed in Annex III (see *above*). Secondly, such systems must entail a risk of harm to health and safety, or a negative impact on fundamental rights, and this risk must be equivalent to or greater than that associated with the high-risk AI systems already listed in Annex III. What clearly emerges from the scope of the regulation at issue here is the

[215] Art. 97 EU Regulation 2024/1689.

reference to the 'risk of harm to health and safety, or of negative impact on fundamental rights': on the one hand, it stands as a parameter of the *ex ante* assessment to be carried out; on the other, it represents one of the main *rationes* that characterise the Regulation as a whole and finds natural application in a context in which the evolution of automated systems may be beyond man's control. For these reasons, the Regulation provides in Article 27 for an 'impact assessment on fundamental rights' to be carried out in the following terms. Prior to the deployment of a high-risk AI system, with the exception of the systems referred to in point 2) of Annex III, *deployers* who are public law bodies, private entities providing public services, as well as *deployers* of high-risk AI systems referred to in Annex III, point 5, letters b)[216] and c)[217] , are obliged to conduct an assessment of the impact on fundamental rights resulting from the use of such systems. This assessment must include the following elements a description of the processes in which the high-risk AI system will be used, in accordance with its intended purpose; a description of the time period during which each AI system will be used and the frequency of its use; identification of the categories of natural persons and groups that could be affected by its use in the specific context; an analysis of the specific risks of harm that could affect the categories of natural persons or groups indicated, taking into account the information provided by the provider in accordance with Article 13[218] ; a description of the human oversight measures envisaged, in line with the instructions for use of the system; and finally, the measures to be taken should the risks materialise, including provisions on internal *governance* and grievance mechanisms. Article 27 further specifies that the *deployer* has the option, in similar cases, to refer to previously conducted fundamental rights impact assessments or existing impact assessments carried out by the provider. If, during the deployment of the high-risk IA system, the *deployer* finds that any of the elements indicated in paragraph 1 have changed or are out of date, the deployer must take the necessary steps to update the information. Once the fundamental rights impact assessment has been completed, the *deployer* is obliged to notify the results of the analysis to the market surveillance authority, attaching the completed template referred to in paragraph 5 as part of the notification. However, in the case provided for in Article 46(1) of the Regulation[219] *, deployers* may be exempted from this notification obligation. Where one of the obligations set out in

[216] Annex III, point 5(b) AI ACT: "AI systems intended for use in assessing the creditworthiness of natural persons or establishing their creditworthiness, with the exception of AI systems used for the purpose of detecting financial fraud".
[217] Annex III(5)(c) AI ACT: "AI systems intended to be used for risk assessment and pricing with respect to natural persons in the case of life insurance and health insurance."
[218] Art. 13 AI ACT.
[219] Art. 46 para. 1 AI ACT.

this Article has already been fulfilled by means of a data protection impact assessment pursuant to Article 35 of Regulation (EU) 2016/679[220] or Article 27 of Directive (EU) 2016/680[221] , the fundamental rights impact assessment referred to in paragraph 1 shall supplement that data protection impact assessment. Finally, the Office for Artificial Intelligence (IA) is tasked with developing a questionnaire template, including through the use of automated tools, in order to facilitate compliance by *deployers* in a simplified manner. With respect to what is envisaged by the regulation, it must be noted that it presents certain critical points. On the one hand, the required assessment operations entail significant economic burdens for companies and the use of considerable resources, since - in view of the particular complexity of the activity, which entails a necessary balancing act between rights and artificial intelligence systems - it obliges providers and users to resort to qualified external experts to perform such assessments. In any case, they would remain responsible for critical decisions related to the use of AI systems. In the absence of sector-specific guidelines, an accurate assessment is not possible, and the lack of such guidelines could lead to overly discretionary self-regulatory practices on the part of obliged parties. On the other hand, the regulation stipulates that for high-risk systems, traceability of their operation must be ensured, ensuring that *the output* is sufficiently transparent to allow proper use by users. Furthermore, it is stipulated that these systems must be developed to offer an adequate level of accuracy, robustness and computer security. They must also incorporate human-machine interface tools that allow for effective human supervision that can prevent and mitigate risks to health, safety and fundamental rights. Such systems, usually deployed in sensitive areas such as medical devices, vehicles, personnel selection processes, critical infrastructures and access to essential services, require a high level of precision and control to ensure user protection and compliance with regulations .[222]

Artificial Intelligence systems with residually identified risk. In envisaging a risk-based approach, Recital 26) of the Regulation[223] refers to transparency obligations for 'certain AI systems',

[220] Article 35 of Regulation (EU) 2016/679.
[221] Article 27 of Directive (EU) 2016/680.
[222] On this point, see Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19. P. 120-121; see also Fabiano, L. (2024). Algorithmic Recommendation and Artificial Intelligence in the US and European experience between strategic needs and fundamental rights protection. IRCOCERVO, 23(1), 440-456.pp.124-125; Novelli, C. (2024). The European Artificial Intelligence Act: some implementation issues. FEDERALISMS. EN, 2, 94-113. Pp. 100 ff.
[223] The text of the Recital states, in particular, 'In order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk-based approach should be used. Such an approach should adapt the typology and content of these rules to the intensity and extent of the risks that may be generated by AI systems. It is therefore necessary to prohibit certain unacceptable AI practices, establish requirements for high-risk AI systems and obligations for relevant operators, as well as transparency obligations for certain AI systems."

which are identified by excluding systems with unacceptable risk and those with high risk[224] . In this case, it would seem necessary to read in conjunction with the aforementioned Recital 26) Article 50 of the Regulation, entitled "Transparency obligations for providers and deployers of certain AI systems". It provides that providers of artificial intelligence systems interacting directly with natural persons are required to ensure that such systems are designed and developed in such a way that users are adequately informed of the nature of the interaction with an AI system, unless such interaction would not be apparent to a reasonably informed and attentive person considering the circumstances and context in which the use takes place. The duty to inform does not extend, however, to AI systems used for the detection, prevention, investigation or prosecution of crime. In addition, providers of AI systems, including those intended for general purposes (GPAI, in the English acronym) that generate audio, video, image, or synthetic textual content, must ensure that the outputs of such systems are marked legibly and detectable as artificial products or manipulations. They must then ensure that their solutions are technically effective, interoperable and reliable, also taking into account the specificities of the various types of content, implementation costs and the state of the art, as well as technical standards. This obligation does not apply, however, in cases where AI systems are used for standard editing assistance or do not substantially alter the input data provided by the *deployer* or the semantics of such data, nor where use is authorised by law for purposes related to criminal offences. Where an AI system

[224] Some doctrine refers to this type of risk as 'low or minimal risk', this is Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19. The Authors, in this regard, state that 'they are identified by difference from unacceptable or high risk systems. Specifically, AI systems involving low or minimal risk are all those lawful systems that are only required to comply with certain minimum transparency requirements. In particular, the Regulation requires the supplier to ensure that the person exposed to an AI system is informed, in a timely, clear and comprehensible manner, that they are interacting with an AI system. In addition, 'where necessary', it must be indicated what functions the system performs, whether there is control over the system by a natural person, the identity of the person responsible for the system, an indication of the exposed person's rights, and the actions that the exposed person can take to object to the use of the AI system. In addition to the duties of providers, the Regulation prescribes specific obligations also for users of high-risk AI systems. Users are, in fact, required to ensure the correctness and quality of the data used by the system. In addition, users must comply with the supplier's instructions but, if they consider that use in accordance with the instructions presents a risk, they must notify the supplier and, if the latter is unavailable, the user must notify the competent control authority of the incident or malfunction." The same authors distinguish a further category of risk, which they refer to as 'AI systems with specific transparency risks' and is defined in these terms as 'systems in relation to which operators have to fulfil substantially less onerous obligations, mainly of an informative nature (e.g. in relation to the use of chatbots, making it clear to users that they are conversing with a machine)'. Other doctrine, in particular citing Gaetana Natale, instead proposes the following classification: 1) unacceptable risk, 2) high risk, 3) limited risk and 4) low and minimal risk. The European Commission (see on this point https://digital-strategy.ec.europa.eu/it/policies/regulatory-framework-ai) also uses the same partitioning. However, if we wish to read the Regulation from the point of view of a reasoned and a posteriori study, it seems that "limited risk" and "minimal risk" have as their common denominator the residual identification of the other two types of risk.

is used for emotion recognition or biometric categorisation, *deployers* are obliged to adequately inform natural persons interacting with such system about its operation, processing personal data in accordance with Regulation (EU) 2016/679, Regulation (EU) 2018/1725 and Directive (EU) 2016/680, where applicable. The obligation to provide information does not apply, however, when such systems are used for the detection, prevention, investigation or prosecution of criminal offences, provided that appropriate protections for the rights and freedoms of third parties are guaranteed and that Union law is respected. Similarly, *deployers* of an AI system that generates or manipulates images or audio or video content that constitutes a 'deep fake' must ensure that the public is informed of the artificial nature of such content. This obligation does not extend when the use of such content is authorised by law for the detection, prevention, investigation or prosecution of crimes. Furthermore, in the event that the content forms part of artistic, creative, satirical or fictitious works, the obligation of transparency is limited to the disclosure of the manipulation or artificial generation of the content without affecting the enjoyment of the work itself. Finally, *deployers* of AI systems that generate or manipulate texts intended to inform the public on matters of public interest are required to disclose that such content has been artificially generated or manipulated. This obligation does not apply when the use is authorised by law for criminal purposes or when the content has undergone human review or editorial control, with a natural or legal person responsible for publication. The above information must be provided to the natural persons concerned in a clear and discernible manner, at least at the time of first interaction with the IA system, and must comply with the applicable accessibility requirements. The transparency provisions are without prejudice to other requirements and obligations under the Regulation itself[225] or other applicable European or national legislation . The category of IA systems whose risk is to be identified on a residual basis may also include those systems referred to in Recital 53) of the Regulation[226] . In that case, the regulation considers 'specific cases in which AI systems relating to predefined areas [...] do not entail a significant risk of affecting the legal interests protected within those areas' In other words, the aforementioned Recital refers to situations in which the use of AI systems as such does not materially influence the decision-making process or affect those interests in a material sense. For the purposes of this Regulation, an artificial intelligence (AI) system that does not materially affect the outcome of the decision-making process is to be understood as a system that has no substantial impact either on the final outcome or on the decisions taken, whether

[225] CHAPTER III AI ACT
[226] Recital 53) AI ACT

human or automated. In this context, an AI system that does not materially affect the decision-making process could arise in several situations, which fulfil one or more of the following conditions. First, the first condition concerns AI systems intended to perform strictly procedural tasks, such as those that transform unstructured data into structured data, that classify documents into specific categories, or that are used to detect duplicates in a large number of applications. Such tasks, limited and restricted in themselves, entail minimal risks that do not increase, even with the use of the system in contexts identified as high-risk. The second condition implies that the task performed by the AI system is aimed at improving the result of a human activity that has already been completed, an activity that could be relevant in high-risk contexts. In these cases, the AI system provides additional value to a pre-existing task, reducing risk and increasing the accuracy of results. An example of this concerns AI systems used to improve the language of previously drafted documents, e.g. by refining the tone, academic style or aligning the text to certain communicative or *brand* standards. The third condition is that the AI system is designed to detect decision-making patterns or deviations from previously adopted patterns. In this case, the risk is limited, since the use of the system follows a pre-existing human assessment, and is not intended to replace that assessment without appropriate human intervention. An example would be an AI system used to check whether a teacher has deviated from a previously established assessment model, with the aim of reporting inconsistencies or anomalies. The fourth condition relates to AI systems intended to perform tasks that are only preparatory to a final assessment involving a high-risk AI system, thus reducing the likelihood that the output of the system itself poses a risk to the subsequent assessment. An example of such a condition concerns intelligent solutions for file management, which include functions such as indexing, searching, processing text or voice data, or linking data to other sources, as well as AI systems used for document translation. In any case, it is important to consider that AI systems used in high-risk contexts, as outlined in the Annex to the Regulation under review, pose significant risks to the health, safety or fundamental rights of individuals, especially where the system involves profiling activities within the meaning of the definitions set out in Article 4(4) of Regulation (EU) 2016/679, Article 3(4) of Directive (EU) 2016/680, or Article 3(5) of Regulation (EU) 2018/1725 (on this point, see more fully *above*). In order to ensure traceability and transparency, a supplier who considers that an IA system does not pose a high risk based on the conditions described above, must draw up documentation regarding the assessment before the system is placed on the market or put into service, and provide this documentation to the competent national authorities upon request. Furthermore, the

supplier is obliged to register the system in the EU database established under this regulation. Finally, in order to provide practical guidelines for the implementation of the conditions under which the IA systems described in the Annex to the Regulation may, exceptionally, not be considered high-risk, the European Commission, after consulting the IA Council, is to publish guidelines defining the practical implementation of these conditions, accompanied by an illustrative list of use cases of high-risk IA systems and use cases that are not.

Artificial Intelligence Systems with Systemic Risk. Article 51 of the Regulation[227] , entitled "Classification of AI models for general purposes with systemic risk", introduces a special level of risk. In particular, it is a risk that the regulation associates with so-called general-purpose AI models. An artificial intelligence (AI) model intended for general purposes is classified as an AI model with systemic risk because the models employed present or may present an inherent risk linked to the high impact capacity of AI models intended for general purposes, which may have a significant effect on the EU market, deriving from the extent of such impact or from the negative effects, both real and reasonably foreseeable, on public health, security, fundamental rights and society as a whole. In addition to these potential impacts, there are actual or foreseeable negative consequences, relating to serious incidents, disruptions in critical sectors and considerable damage to public health and safety, as well as the dissemination of illegal, false or discriminatory content, which could lead to harmful bias and discrimination, with tangible risks to individuals, communities and society. Risks associated with general-purpose AI also include those related to intellectual property rights[228] , plus the need to respect and protect confidential business information and business secrets, in accordance with EU and national law[229] . Risks to public and economic security, dissemination of disinformation, damage to privacy, and threats to democratic values and human rights are also relevant[230] . In this context, it is crucial to pay attention to risks arising

[227] The text of the provision reads: "A general purpose AI model is classified as a general purpose AI model with systemic risk if it meets one of the following conditions

a)it has high impact capabilities assessed on the basis of appropriate technical tools and methodologies, including indicators and benchmarks;

b)on the basis of a decision of the Commission, ex officio or following a qualified opinion of the Scientific Panel, it has equivalent capacity or impact to that referred to in point (a), taking into account the criteria set out in Annex XIII.

2. A general purpose AI model shall be presumed to have high impact capability in accordance with paragraph 1(a) when the cumulative amount of computation used for its training measured in floating point operations is greater than 10^{25}.

3. The Commission shall adopt delegated acts in accordance with Article 97 to amend the thresholds referred to in paragraphs 1 and 2 of this Article, as well as to supplement benchmarks and indicators in the light of evolving technological developments, such as algorithmic improvements or increased hardware efficiency, where necessary, so that those thresholds reflect the state of the art."

[228] See on this point Recital 48) of the AI Act.

[229] Consider, more extensively on this point, the wording of Section 53 AI Act

[230] On this point, read Recital 110) AI ACT.

from potential intentional misuse, or unintentional control problems relating to alignment with human intent. These risks include chemical, biological, radiological and nuclear risks, such as reducing barriers to entry that could facilitate weapons development, project acquisition or misuse, as well as the use of offensive cyber capabilities, such as the detection, exploitation or operational use of vulnerabilities. In addition, the risks associated with the interaction and use of tools that can compromise the ability to control physical systems and interfere with critical infrastructure are highlighted. It is also emphasised that the development or use of AI models may entail the risk of serious incidents, and cybersecurity protection related to systemic risks arising from malicious use or attacks must take into account issues such as accidental leakage of models, unauthorised releases, circumvention of security measures, as well as defence against cyber attacks, unauthorised access or theft of models[231] . Finally, systemic risks may also emerge from the design, operation and use of large online platforms, as well as online search engines, where the design of the algorithmic systems used in the service may contribute to such risks .[232]

What emerges from the normative and jurisprudential architecture of EU Regulation 2024/1698 is, first and foremost, a *focus* not so much on the individual as on Artificial Intelligence systems.

3. *Large Language Models*. A test case for the European legislator.

As illustrated, in the context of the rapid evolution of artificial intelligence and the growing need for regulation, the European Union was faced with the imperative of developing an advanced legal framework to govern and regulate related technological innovations. The European Commission's proposal for an Artificial Intelligence Regulation (AI Act), published in 2021, was a significant step towards a harmonised regulatory framework for AI in Europe. However, it is important to note that the publication of this proposal took place before the widespread emergence of generative AI. As a result, the Commission's proposal had not provided a clear definition of generative AI, nor specifically addressed the risks associated with this different form of AI. It was only following the rapid spread of ChatGPT, and of the various *Large Language* Models in general, that discussions among EU legislators

[231] Recital 115) AI ACT.

[232] Recital 118) AI ACT. For this reading of systemic risk, see more extensively Corsi, R., & D'Albergo, E. (2024). The politics of general purpose artificial intelligence: socio-technical imaginaries, democracy and policy frames in European regulatory decision-making ("EU AI ACT"-2022-2024). IM@ GO, 23, 109-130. Pp. 117-118

focused on whether the proposal should be amended to include specific provisions for generative AI, as well as on the regulation of new categories of AI, such as *Foundation Models* and General Purpose Artificial Intelligence (GPAI). This marked a radical transformation from the traditional categories initially considered by the Commission. Moreover, in the already analysed economics *of the* risk-based model, the provision of an *ad hoc* risk 'layer'

3.1. *Foundation Models* in the proposed Artificial Intelligence Regulation.

In the long journey that led to the current set-up, the debate around generative AI, which led to very recent amendments in the most recent text version, was decisive. As analysed above, the AI Act is based on a *risk-based* model, whereby a distinction is made between different levels of risk. Certain types of AI applications are considered to entail unacceptable risks for individuals and are therefore prohibited under the Act. As regards generative AI, several risks are associated with them, not least the dissemination of 'toxic' content or false information. The proposal text that was under consideration by the European institutions before the current final version, as far as its subject matter is concerned, contained a statement in Article 1(1) that one of the aims of the AI law was to ensure a high level of protection of health, safety and fundamental rights. The definition of an AI system in Article 3(1) was then amended to align it more closely with the work of international organisations working on artificial intelligence, in particular the OECD. In addition, the corresponding Recital 6) further detailed the key features of the definition and clarified that the definition is not intended to cover traditional software systems or simpler programming approaches that rely on rules defined solely by natural persons to perform operations automatically. Furthermore, the Commission was mandated to develop guidelines on the application of the definition of an AI system. In line with the revised mandate of the Council, in Article 29a, the text included a simplified version of the obligation for certain operators to carry out an IA. The obligation concerned only operators who are bodies governed by public law, private actors providing public services and operators providing banking and insurance services using AI systems listed as high-risk. The fundamental rights impact assessment was to be carried out only for aspects not covered by other legal obligations, such as the data protection impact assessment under the GDPR, and procedurally aligned with existing processes in order to eliminate possible overlaps and additional burdens. Subsequently, the agreement on which an initial compromise was reached included new provisions on general purpose artificial intelligence (GPAI) models in Articles 52a - 52e. These new rules introduced horizontal obligations for

all GPAI models, including keeping them up-to-date and making technical documentation available to the IA Bureau and competent national authorities upon request. They also included the provision of certain information and documentation to downstream suppliers for the purposes of compliance with the IA Act. There were also some additional requirements in this version for models with systemic risks, including performing a model assessment, assessing risk and taking risk mitigation measures, ensuring an adequate level of cybersecurity protection, and reporting serious incidents to the IA Bureau and the competent national authorities. Compliance with these requirements could be achieved through codes of conduct, to be developed by industry, with the participation of the Member States (through the AI Board) and facilitated by the AI Office. The development of codes of conduct was to be an open process to which all stakeholders, both companies and civil society and academia, would be invited. The AI Office would also evaluate these codes of conduct and approve them formally, or, if they were inadequate to cover the obligations, provide common rules for the implementation of the obligations through an implementing act. The compromise agreement also provided that the standards would be a possibility for compliance in the future. The classification of GPAI models as posing systemic risks initially depended on the capacity, either on the basis of a quantitative threshold of the cumulative amount of calculation used for training measured in floating point operations (FLOP), or on an individual designation decision of the Commission taking into account the criteria listed in Annex IXc (e.g. the number of parameters, the quality and size of the dataset, the input and output modalities, or the scope measures in business users). The initial FLOP threshold for this was set at 10^25, thus higher than initially desired by the EP, but lower than the Council mandate, which was 10^26. However, in this context, it is important to note that the Commission had been - and is also, according to the letter of the finally approved version - obliged to adapt the threshold in the light of evolving technological developments, such as algorithmic improvements or increased hardware efficiency, in order to keep up with the state of the art, reflecting the capabilities recorded in the most advanced general artificial intelligence models. As far as copyright is concerned, the compromise agreement stipulated the implementation by GPAI model providers of a policy to comply with Union copyright law, as well as to make publicly available a sufficiently detailed summary of the content used for training the general purpose AI model, based on a template provided by the AI Office. Recital 60k explained that this summary should not be technically detailed but exhaustive at a general level, taking into account the need to protect trade secrets and confidential business information. Finally, with regard to GPAI systems, the previous compromise agreement

defined them as artificial intelligence systems based on GPAI models that can serve a variety of purposes. Article 28 included rules to align them with the risk-based approach of the AI Act. Suppliers of such systems were required to provide all information and elements to downstream suppliers of high-risk AI systems so that they could comply with the respective requirements, including for compliance assessment purposes. Article 63a further clarified the conditions under which national market surveillance authorities should cooperate with the IA Bureau to ensure compliance of GPAI systems that could have been used directly by operators for at least one purpose classified as high-risk, when they consider GPAI the systems are not compliant.

With specific reference to Generative Artificial Intelligence, reading the version of the text of the AI Act as of 13 March 2024 appears decisive. Indeed, at Recital 60c), large generative AI models were considered a typical example of a generic AI model, since they allow the flexible generation of content (such as in the form of text, audio, images or video) that can easily accommodate a wide range of distinctive tasks. Again, in Recital 60(i), we could read about how general purpose models, particularly large generative models capable of generating text, images and other content, presented unique opportunities for innovation but also challenges for artists, authors and other creators and the way their creative content is created, distributed, used and consumed. The development and formation of such models requires access to large amounts of text, images, video and other data. *Text* and *data mining* techniques can be widely used in this context to retrieve and analyse such content, which may be protected by copyright and related rights. Any use of copyright protected content requires the permission of the relevant rights holder, unless copyright exceptions and limitations apply. Directive (EU) 2019/790 introduced exceptions and limitations allowing reproductions and extractions of works or other material, for text and data extraction purposes, under certain conditions. Under these rules, right holders could choose to reserve their rights to their works or other material to prevent the extraction of text and data, unless this was for scientific research purposes. If the *opt-out* right had been expressly reserved in an appropriate manner, providers of general purpose AI models would have had to obtain permission from the right holders if they wanted to perform *text* and *data mining* on such works. In fact, whereas in the previous version the more general *Foundation Models* were taken into account, in the text now under consideration the *Large Language Models* were also taken into account more specifically. In particular, Annex IX (c) (d) refers to the input and output modalities of the model, such as text-text (*Large Language Models*), text-image, multimodality and the state-of-the-art thresholds for determining high-impact capabilities for each modality

and the specific types of input and output (e.g. biological sequences), for the purposes of determining systemic risk under Article 52 in the version under review. This *excursus* allows some considerations to be made. The so-called *Foundation Models*, including GPAI and generative AI, in its most widespread form as *chatbots*, were initially considered as low or no risk technologies in the first version of the regulation (EU Commission, 2021). However, a significant change of perspective occurred in November 2022, when, at the same time as the launch of ChatGPT by OpenAI and in view of the start of the trilateral negotiations, the Council of the European Union proposed to introduce specific regulation for generative AI models, while maintaining a less stringent discipline than the one subsequently adopted, in order to avoid conflicts between the different political visions of the various member states. The Parliament's negotiating position, adopted in June 2023, introduced far more stringent measures, as, based on a deeper understanding of the risks associated with the use of fundamental models, it recognised the need to develop them, but at the same time called for mitigation of 'foreseeable risks to health, security, fundamental rights, the environment, democracy and the rule of law'. In November 2023, the joint declaration of Italy, France and Germany on the obligation to self-regulate GPAI systems threatened to undermine the agreements, as the governments of these states had opposed the regulation of the core models, influenced by lobbying by the French start-up Mistral, supported by the former French Secretary of State for Digital, Cédric O, and the leading German AI company, Aleph Alpha. These actors, as well as many other companies, believed that regulation was incompatible with innovation. This position raised concerns and reactions from various actors, including European start-ups, civil society organisations (NGOs) and expert communities, who were in favour of regulation of fundamental models for the protection of people's rights. Concerns raised in this context related to 'profound risks to society and humanity' that could not be addressed through corporate decisions alone, without the involvement of democratically elected governments. Among the main risks highlighted were: the possibility of flooding communication with content that conveys propaganda and falsehoods, the destabilisation of labour markets through the automation of rewarding jobs, the amplification of discrimination and prejudice, with the further marginalisation of disadvantaged communities and alternative views, the loss of power of political institutions in favour of concentrated power in the hands of a few private companies, and the obsolescence of human beings, with the consequent loss of control over our civilisations. These concerns have had a significant impact on the subsequent stages of the political process, culminating in the approval, in spring 2024, of the Regulation, which, composed of

legal rules and recitals of a discursive nature, makes explicit the rationale and interpretation of the regulatory provisions. In this context, the Regulation defines the framework for regulatory intervention relating exclusively to the component of our interest: the risk representations associated with the use of fundamental models and GPAI, as well as the associated value and belief systems .[233]

3.2. The text of the current Regulation *: General-purpose AI Models* and *General-purpose AI Systems.*

Against this background, the current text of the Regulation has taken into account the very rapid evolution of Generative Artificial Intelligence. With reference to so-called GPAI, the Regulation contains a number of rules. First of all, for their definition, we have to look at Article 3 of the Regulation: in paragraph 66) "AI system for general purposes" is identified as an artificial intelligence system, based on an AI model for general purposes, capable of pursuing different purposes, both for direct use and for integration in other AI systems; it represents, therefore, a type of system that possesses a functional flexibility that allows adaptation to multiple application contexts. In contrast, general-purpose AI models are characterised by significant generality, in that they are capable of competently performing a wide range of distinct tasks and, at the same time, of being integrated into various *downstream* systems or applications. The performance of an AI model, in fact, is a key indicator of its generality, and where a model possesses more than one billion parameters, that model is presumed to demonstrate 'significant generality', as stated in Recital 98), although the definition of parameter may vary depending on the method adopted by the respective AI model. In general terms, parameters are the values used by AI models to infer results from input data. GPAI models are also commonly known as *Foundation Models*, as these models can serve as the basis for numerous *downstream* AI applications, i.e. for a variety of subsequent uses. For instance, large generative AI models, such as *Large Language Models* (LLM), are generally considered GPAI models. However, AI models that have been trained to operate in narrow, defined domains, such as pharmaceutical research or meteorology, often do not fall into the category of GPAI models, precisely because of their limited applicability. AI models that do not meet the requirements of GPAI models can be regulated indirectly, as

[233] On this point see, more extensively, Corsi, R., & D'Albergo, E. (2024). The politics of general purpose artificial intelligence: socio-technical imaginaries, democracy and policy frames in European regulatory decision-making ("EU AI ACT"-2022-2024). IM@ GO, 23, 109-130.

components of domain-specific AI systems. It is important to emphasise that GPAI models differ from GPAI systems, which are AI systems that are based precisely on GPAI models and that, in turn, are capable of serving a variety of purposes. Typical examples of GPAI systems include applications such as *ChatGPT* and Midjourney, which are based on GPT and Midjourney AI models. In this context, suppliers of GPAI models are required to provide the necessary information so that downstream suppliers of AI systems can fulfil their obligations under the AI Regulation. In addition, for those GPAI models that present systemic risks, more stringent obligations are laid down in order to mitigate such risks (see *above*)[234] . Artificial intelligence (AI) models can be used in a wide range of use cases, but in themselves they do not constitute AI systems, as clearly emphasised in Recital 97. Consequently, they cannot be classified as high-risk models under the AI Act. However, if they undergo further refinement or are supplemented with user interfaces, they could be used within AI systems that might instead fall into the high-risk category. The AI Act distinguishes between two categories of GPAI models, depending on the associated risk, that are subject to regulation: GPAI models with systemic risks and 'regular' GPAI models, i.e. without systemic risks. In principle, all GPAI models are obliged to comply with documentation and information requirements, including the obligation to make the data used for training the model accessible, and to implement systems to protect the copyrights of third parties. GPAI models involving systemic risks are subject to additional obligations aimed at improving transparency, recognising potential risks, preventing damage and enhancing cybersecurity. These models are considered to be more dangerous due to their breadth and capacity, as they may lead to misuse or problems with the reliability of the model itself, with potential harmful consequences, such as discrimination, misinformation or compromising the security of AI systems, as mentioned in recital 110. For these reasons, GPAI models with systemic risks must also be reported to the European Commission. In contrast, free and open source GPAI models are considered less risky, precisely because of their inherent transparency, being particularly relevant for research and development, as emphasised in recital 102. For this reason, certain facilities are applied to such models, provided they do not present systemic risks. The requirements for GPAI models do not apply where a model, which does not entail systemic risks, is used exclusively for internal processes, which are not essential for the supply of a product or service to third parties and which do not involve the rights of natural persons, as set out in Recital 97. However, this exemption does not extend to GPAI models that

[234] Voigt, P., & Hullen, N. (2024). Which Requirements Apply to GPAI Models?. In The EU AI Act: Answers to Frequently Asked Questions (pp. 129-149). Berlin, Heidelberg: Springer Berlin Heidelberg. P. 9

present systemic risks. In any event, the implementation of the requirements regarding GPAI models must always be appropriate and proportionate, as highlighted in Recital 109[235] . The provider of a general purpose artificial intelligence (AI) model (GPAI) is identified as any entity that develops such a model or engages others in the development, subsequently placing the model on the market under its own name or trademark, as provided for in Article 3(3). GPAI models may be placed on the market by different means, such as software libraries, application programming interfaces (APIs), direct downloads or physical copies, as referred to in recital 97. In this context, suppliers of GPAI models play a crucial role in the artificial intelligence value chain, as emphasised in recital 101, since downstream suppliers wishing to integrate such models into their AI systems need to have a thorough understanding of the technical characteristics of GPAI models in order to properly fulfil their obligations, including those relating to high-risk AI systems.In particular, providers of GPAI models are required to comply with four main obligations under Article 53(1) they must prepare and keep up-to-date technical documentation of the model, which includes the training process, testing and evaluation results, making this documentation available to the IA Office and the competent national authorities upon request; they must prepare, update and make available information and documentation to IA system providers who wish to integrate the GPAI model into their systems; they must implement a strategy for copyright compliance; and finally, they must prepare and publish a detailed summary of the content used for training the model, according to the template provided by the IA Office. For GPAI models placed on the market under a free and open source licence, certain simplifications are envisaged, provided that such models do not present systemic risks. The technical documentation of a GPAI model, which must be provided upon request by the IA Office and the competent national authorities, must include at least the elements provided for in Annex XI. However, providers of free and open source licensed GPAI models, which do not pose systemic risks, are exempted from the obligation to prepare detailed technical documentation, provided they comply with the relevant requirements. Annex XI specifies the general information to be provided by all providers and the additional information required for models with inherent systemic risk. As a minimum, the general technical documentation must contain a general description of the GPAI model, including the tasks the model is capable of performing, the type of IA systems with which it can be integrated, the applicable policies for acceptable use, the release date, the deployment methods, the architecture and number of parameters, the

[235] Voigt, P., & Hullen, N. (2024). Which Requirements Apply to GPAI Models?. In The EU AI Act: Answers to Frequently Asked Questions (pp. 129-149). Berlin, Heidelberg: Springer Berlin Heidelberg. Pp. 129-130.

input/output mode and format, and the licence under which the model is offered. In addition, a detailed description of the constituent elements of the model and information pertinent to the development process must be included, including the technical specifications necessary for integrating the model into AI systems, design and methodological choices, and information on the data used for training, testing and validation, including the type, source and characteristics of the data, along with measures taken to detect any biases or inadequacies in the data sources. Information must also be provided on the computational resources used, as well as the known or estimated energy consumption of the model. The European Commission is empowered to amend the list of elements required for the technical documentation of a GPAI model by means of a delegated act, pursuant to Article 53(6). With regard to GPAI models that present systemic risks, providers must provide additional information in the technical documentation. In any case, providers of free and open source licensed GPAI models that do not present systemic risks are exempt from the obligation to provide technical documentation, provided they comply with the relevant requirements[236] . Suppliers of general-purpose AI (GPAI) models are obliged to make the necessary information available to downstream suppliers wishing to integrate such models into their AI systems, as stipulated in Article 3(68). This information must enable downstream providers to implement the GPAI model in their IA system in a professional, secure and compliant manner, particularly with regard to high-risk IA systems. The information provided must enable downstream providers to gain an adequate understanding of the capabilities and limitations of the GPAI model and to properly fulfil their obligations under the AI Act. However, providers of GPAI models that are made available under a free and open source licence, provided they do not pose a systemic risk, are exempt from the obligation to provide such information to downstream providers, provided they comply with the relevant requirements. In accordance with Annex XII, the supplier of a GPAI model must make available the following information, which in part reflects the elements to be included in the technical documentation of the model. First, a general description of the GPAI model must be provided, including the tasks the model is capable of performing, the type and nature of AI systems with which it may be integrated, the applicable policies for acceptable use, the release date, and the methods of deployment. Where relevant, details of how the GPAI model interacts with hardware or software that is not part of the model itself, or how it can be used for such interaction, must also be specified. The versions of related

[236] Voigt, P., & Hullen, N. (2024). Which Requirements Apply to GPAI Models?. In The EU AI Act: Answers to Frequently Asked Questions (pp. 129-149). Berlin, Heidelberg: Springer Berlin Heidelberg. Pp. 131-133.

software that may be required for the GPAI model to function (e.g. whether the GPAI model is only compatible with a specific version of an operating software) must also be indicated. Other information to be provided includes the architecture and number of parameters of the model, the mode (text, image, etc.) and format of inputs and outputs, and the licence of the model. In addition, a detailed description of the elements of the GPAI model and the development process must be provided, including the technical means required to integrate the model into AI systems, such as instructions for use, infrastructure and required tools. The mode and format of the inputs and outputs, as well as their maximum size (e.g. the length of the context window or the number of tokens that the model can process), must be clearly specified. If applicable, information on the data used for training, testing and validating the model must be provided, including the type and origin of the data, as well as the curation methods adopted (such as data cleaning or filtering). It is important to emphasise that information on the mode and format of inputs and outputs should be included in both the general description of the GPAI model and the more detailed description of its components and development process, with a possible overlap of content between the two sections, e.g. regarding the mode of input and output. While the general description of the model highlights the main features, the description of the components and the development process must provide more specific details, such as the maximum size of the output. Finally, suppliers of GPAI models with a free and open source licence that do not present systemic risks are exempted from the obligation to provide information to downstream suppliers, in accordance with the relevant requirements[237] . With regard to information on the data used for training GPAI models, GPAI model providers are required to ensure transparency regarding the datasets used for model training, in order to enable downstream providers integrating the GPAI model into their AI systems to more accurately assess the risks and legality of the data used. This includes, for instance, compliance with copyright and data protection legislation, as well as an assessment of the performance they can reasonably expect from the AI model. As a crucial element of an AI system, the AI model is closely linked to the quality and management of the data feeding it. For providers of high-risk AI systems, in particular, training data information is essential to fulfil their data governance and risk management obligations to ensure compliance with applicable regulations. Therefore, GPAI model providers are obliged to prepare and make publicly available a sufficiently detailed summary of the data used for training the model, following a template provided by the Office

[237] Voigt, P., & Hullen, N. (2024). Which Requirements Apply to GPAI Models?. In The EU AI Act: Answers to Frequently Asked Questions (pp. 129-149). Berlin, Heidelberg: Springer Berlin Heidelberg. Pp. 133-134.

of IA. In this process, it is crucial that business secrets and confidential information are not made public, as they must be adequately protected, as stipulated in recital 107. The summary, while not having to be technically detailed, must be comprehensive enough to provide the information necessary for all parties with legitimate interests (e.g., rights holders) to exercise and enforce their rights under EU law. For instance, the list of main data collections, such as public datasets or archival content that is not publicly accessible, may be included in the summary. Overall, the summary should constitute a 'narrative explanation' that provides a clear overview of the use of protected works, helping copyright holders understand how their works may be used. To this end, the IA Office is mandated to provide a template that is 'simple and effective' for drafting this summary, as mentioned in recital 107, so that the information is easily understandable and useful for all parties involved[238] . Providers of GPAI models are required to cooperate with the competent national authorities and the European Commission in exercising their supervisory functions over AI systems and GPAI models in accordance with the AI Act. The need for cooperation depends on the potential risk of the AI model and the principle of proportionality. This obligation to cooperate is reinforced by specific reporting requirements for providers of GPAI models that pose a systemic risk. Systemic risk refers to models that, due to their capabilities or impact, may have significant adverse effects on public health, security, fundamental rights or society in general, and may propagate along the value chain. A GPAI model is considered systemic risk when meets certain conditions (Article 51(1)). These models must have high-impact capabilities, which must be assessed with appropriate technical tools and methodologies, such as indicators and benchmarks. In addition, models may be designated as systemic risk by the European Commission, taking into account specific criteria set out in Annex XIII. The Commission may decide to designate a model as high-impact on its own initiative or following qualified advice from a scientific panel, if the model has capabilities or impacts comparable to those of a high-impact model. These capabilities mainly refer to advanced GPAI models, referred to as 'frontier models', although this terminology is not used in the text of the AI Act. A GPAI model is presumed to have high-impact capabilities if the cumulative amount of floating-point operations used in its training exceeds a specific value of 10^25 operations (Section 51(2)). Floating-point operations include all mathematical operations or assignments involving floating-point numbers, which are a representation of real numbers on computers. This amount of calculation should not be confused with the term FLOPS, which indicates

[238] Voigt, P., & Hullen, N. (2024). Which Requirements Apply to GPAI Models?. In The EU AI Act: Answers to Frequently Asked Questions (pp. 129-149). Berlin, Heidelberg: Springer Berlin Heidelberg. P.137.

the processing speed in operations per second. The European Commission has the possibility to establish further indicators and parameters which, if exceeded, lead to a presumption of high impact capability (Article 51(3)). If a GPAI model is classified as having high impact capability, it will be considered a systemic risk model and will go through a formal classification procedure. At this stage, the model provider may challenge the presumption of systemic risk by proving that, despite the high-impact capabilities, the model does not actually pose a systemic risk. Alternatively, the European Commission has the power to classify a Generative Artificial Intelligence (GPAI) model as posing a systemic risk. In order to determine whether a given model should qualify as a GPAI model with systemic risk, as set forth in Annex XIII, it is necessary to examine a number of indicators, which include, but are not limited to, the following the total number of parameters of the model, reflecting its complexity, the quality or size of the dataset employed, which may be measured, for example, by the number of tokens, the amount of computational power expended in training the model, which may be quantified in floating point operations or, alternatively, represented by a combination of other variables, such as the estimated cost of training the time required to complete it or the expected energy consumption, the model's input and output modes, which may include different types, such as text-to-text conversion, typical of language models (LLM), text-to-image transformation, characteristic of image-generative models, multimodality, and the definition of state-of-the-art thresholds to determine the high-impact capabilities for each mode; This also includes the identification of the specific type of input and output, as in the case of artificial intelligence models capable of generating images or biological sequences, such as DNA, to be used, for example, in a medical context, the assessment of benchmarks and model capabilities, also taking into consideration the number of tasks that the model can perform without requiring additional training, its adaptability to new and different tasks, as well as the degree of autonomy and scalability it possesses, together with the tools at its disposal, and finally, the number of end-users, both commercial and non-commercial, who have registered the use of the model within the European Union. Even if a GPAI model has been trained with less than 10^{25}floating point transactions, the European Commission may still classify it as a systemically risky GPAI model, provided that the classification criteria are met and, in addition, there is a comparable risk that justifies such a classification. In addition to the general provisions governing all providers of GPAI models, those operating GPAI models involving systemic risks are required to be informed of all associated systemic risks. Such providers must, in fact, also fulfil the specific obligations set out in Article 55(1), which include the obligation to assess the GPAI model in accordance

with standardised protocols and tools that reflect the state of the art, including the implementation and documentation of adversarial tests, such as red teaming, aimed at identifying and mitigating systemic risks; the need to assess and address potential systemic risks at the EU level, considering the relevant sources, that could arise from the development, marketing or use of the GPAI model with systemic risk. In addition, suppliers are obliged to promptly inform the IA Office and, when appropriate, the competent national authorities of any serious incidents and possible corrective actions to be taken to resolve such issues. In addition to the general requirements laid down, the technical documentation of a GPAI model classified as systemic risk must include, in addition, a number of specific elements, as laid down in Article 53(1)(a), Section 2 of Annex XI. In particular, it must contain: a detailed description of the assessment strategies, including the results obtained, based on publicly available assessment protocols and tools or other assessment methodologies adopted. The evaluation strategies should also include the criteria and metrics used to perform the evaluation, as well as the methodology used to identify limitations in system performance; if applicable, a detailed description of the steps taken to conduct adversarial testing, both internal and external (such as red teaming), as well as any model adaptations, including alignment and tuning; and, if applicable, a detailed description of the system architecture, explaining how the various software components interact with each other, feeding into and integrating into the overall processing. In addition, providers of GPAI models with systemic risk may, where necessary, make use of harmonised standards and codes of conduct to demonstrate their compliance with the additional requirements for providers of AI models with systemic risk, as set out in Article 55(2). Providers of GPAI models are required to notify the European Commission without delay if their model meets the conditions that give it a high impact capability, thereby qualifying it as a systemic risk GPAI model, as set out in Article 52(1). The notification must be sent within two weeks from the moment when the conditions have been met or, in any case, from the moment when the provider is aware of the likelihood of those conditions being met. In this notification, the supplier is also obliged to provide the information necessary to demonstrate the high-impact capability of the GPAI model. If the European Commission does not receive such a notification, it may proceed without delay to designate the GPAI model as having a systemic risk. In the event that the providers, when informing the European Commission about the high-impact capabilities of the model, consider that, despite these capabilities, the model does not present an exceptional systemic risk, they may present sufficiently substantiated arguments to justify this position, as provided for in Article 52(2). In such a case, the European Commission is obliged to

examine the arguments provided by the provider. If the Commission does not agree with those arguments and considers that the GPAI model should nevertheless be classified as systemic risk, it shall reject the provider's arguments and the model shall be deemed to be systemic risk, in accordance with Article 52(3). The classification of a GPAI model as systemic risk is not necessarily subject to the notification of the provider, as the European Commission may, on its own initiative or following a qualified warning from the Panel, designate a model as systemic risk, in accordance with Article 52(4). Such designation must be based on an assessment of the model, carried out in accordance with the criteria set out in Annex XIII, which may be updated by the European Commission from time to time, in line with the evolution of the state of the art, as provided for in Section 5.2.1. If a GPAI model is designated as systemic risk, the provider has the right to request a reassessment of the model by the Commission, based on objective, detailed and new reasons that have arisen since the designation decision, as set out in Article 52(5). Although the IA Act does not explicitly provide for this, this procedure also applies to reasons that the European Commission did not take into account during the initial assessment before designating the model as systemic risk. The request for reassessment may be submitted no earlier than six months after the initial decision, unless the European Commission failed to consider significant reasons in its assessment. If the Commission changes the assessment criteria, it is obliged to check whether the previous classifications are still justified and, if so, to downgrade them if the models are no longer found to be systemic risk following the updated criteria. Finally, the European Commission is obliged to publish and keep up-to-date a list of GPAI models identified as posing systemic risk, in accordance with Article 52(6). Codes of conduct, which materialise the requirements of the AI Act for providers of GPAI models, are of great practical importance to ensure compliance with the provisions of this regulation. Although the adoption of such codes is not mandatory, it significantly facilitates compliance with legal obligations. Indeed, GPAI model providers can more easily demonstrate compliance with their obligations if they apply the approved codes of conduct, which is particularly relevant for systemically risky GPAI model providers. These providers can use the approved codes of conduct to prove the compliance of their GPAI models, until the harmonised standards are published. Although harmonised standards and codes of conduct fulfil a similar function, they differ mainly in the way they are drafted: whereas codes of conduct are developed by stakeholders, in particular by the providers of GPAI models themselves, harmonised standards are drafted by the European Commission and European standardisation organisations. Despite this difference, both modes pursue the same objective. Moreover,

compliance with codes of conduct or harmonised standards facilitates the conformity assessment of suppliers of high-risk AI systems that integrate GPAI models into their systems, making the verification process more efficient. Suppliers that do not apply codes of conduct or harmonised standards still have the possibility of demonstrating compliance with the requirements through alternative procedures, which will have to be assessed by the European Commission, pursuant to Articles 53(4) and 55(2). The AI Office has the task of encouraging and facilitating the development of such codes of conduct, also taking into account international approaches, in order to promote the correct application of the AI Act (Article 56(1)). The codes of conduct are to be made available by 2 May 2025; should it not be possible to complete this process by that date, the European Commission may issue detailed specifications by means of an implementing act, as stipulated in Article 56(9). In addition, the IA Office and the IA Council are responsible for ensuring that the codes of conduct cover at least the general requirements for GPAI model providers (Article 53) and the specific requirements relating to systemically risky GPAI models (Article 55), also taking into account the aspects identified in Article 56(2), including: the methods for ensuring that the information provided by the providers is up-to-date in relation to technological and market developments; the appropriate level of detail to summarise the content used for training; the identification of systemic risks, where relevant, and their causes; and the measures, procedures and arrangements for assessing and managing systemic risks, including related documentation, which must be proportionate to the risks, taking into account the severity and likelihood of the risks. The AI Office may involve GPAI model providers, competent national authorities and other stakeholders, such as civil society organisations, industries, academic bodies and independent experts, in the process of developing codes of conduct (Article 56(4)). The purpose of this involvement is to ensure that the needs and interests of all parties involved, including stakeholders, are taken into account. The codes of conduct must clearly set out objectives, commitments and specific measures, possibly supported by key performance indicators (KPIs), not only to facilitate the monitoring of the objectives by the suppliers, but also to regularly inform the AI Office on the implementation of the commitments through the reporting of these indicators (Article 56.5). The AI Office and the AI Council are in charge of periodically monitoring and evaluating the fulfilment of the objectives contained in the codes of conduct, ensuring that their input fosters the proper implementation of the Regulation. Suppliers not established in the EU are required to appoint an authorised representative based in the EU before placing a GPAI model on the market, as provided for in Article 54(1) (see Section 1.3.10). The authorised representative

must, on the basis of a written mandate, assume the responsibilities of the supplier, acting as its interlocutor for the AI Office and the competent authorities in all matters relating to compliance with the AI legislation (Article 54(4)). However, providers of free and open source GPAI models, which do not pose systemic risks, are exempted from the obligation to appoint an authorised representative (Article 54(6); see Section 5.3). The mandate conferred on the authorised representative must at least provide that the latter may: (i) verify that the provider has fulfilled all obligations under the IA Act (see Section 5.1), (ii) make available to the IA Office and the competent national authorities a copy of the technical documentation as well as the contact details of the provider of the GPAI model (see Section 5.1.1), (iii) provide, upon reasoned request, all information and documentation necessary to demonstrate compliance with the requirements for GPAI templates, and (iv) cooperate with the IA Office and the competent authorities in any action taken with regard to the GPAI template or an AI system based on it, if so requested. If the authorised representative considers that the supplier is acting contrary to its obligations under the AI Act, it must terminate the contractual relationship. In such a case, the authorised representative must immediately inform the IA Office of the termination of the mandate and provide the reasons underlying such decision, as set out in Article 54(5). A supplier of an artificial intelligence (AI) system that integrates one or more GPAI models is considered a downstream supplier, as defined in Article 3(68), and follows, in the AI value chain, the supplier of the GPAI model itself. The obligations of a downstream supplier of an IA system, which may also be the supplier of the IA model, depend on the classification of the IA system according to the risk-based approach established by the IA Act. In particular, the AI system must not be prohibited, as stipulated in Article 5. In the event that the IA system is classified as high-risk, pursuant to Article 6, the downstream supplier is required to comply with all obligations for suppliers of high-risk IA systems, as set out in Article 16. If, on the other hand, the IA system is not classified as high-risk, the downstream provider may only have to comply with the transparency obligations under the IA Act. Providers of GPAI models are obliged to provide downstream providers with sufficient information and documentation regarding the GPAI model, in particular when dealing with models that pose systemic risks, in order to ensure that downstream providers can properly fulfil their obligations under the AI Act if their AI system depends on a GPAI model (recital 101). In addition, downstream suppliers have the right to submit a reasoned complaint to the IA Office if they consider that the GPAI model integrated into their IA system violates the provisions of the AI Act, as provided for in Article 89(2). The AI Act does not impose any specific obligations on distributors, importers and

distributors of GPAI models, as it applies only to distributors, importers and distributors of IA systems, as stated in Articles 2(1)(b) and (d), and 3(4), (6) and (7). Consequently, the IA Act does not directly cover GPAI models, but only IA systems. However, there is no protection gap in this context, as providers of GPAI models, or their authorised representatives, can always be held responsible for complying with the requirements of the AI Act for GPAI models. Indeed, the 'operation' of an AI model always takes place within an AI system, for which there may be specific obligations regarding distributors, importers and dealers, especially in the case of high-risk AI systems . [239]

4. *Large language models,* profiling, personalisation, vulnerability.

The Regulation on Artificial Intelligence (AI) does not constitute an autonomous piece of legislation, but is part of a broader and more articulated European legal context, which provides for coordination with other disciplines proposed by the European Commission, relating to data and digital transformation, in order to guarantee legal certainty. The complexity of this coordination increases if one considers the close interrelation with the General Data Protection Regulation (GDPR), as AI systems, in order to achieve high performance, require huge amounts of data, the availability of which directly influences the accuracy of the predictions produced. Consequently, one of the crucial aspects for the proper functioning of an AI system is the data collection phase. However, the AI Regulation does not deal with the collection and processing of data, which continue to be governed by existing legislation, in particular the GDPR. After data collection, it is necessary to understand how data are used and processed within the AI system. The choice of data sources, in fact, becomes a central element linking all European digital regulatory acts, in line with the goal of creating a governance of innovation based on the principle of 'one continent, one law'. Another common aspect between the GDPR and the AI Regulation is the adoption of the so-called 'risk-based approach', which is based on an ascending pyramid of gravity, where organisations are called upon to demonstrate their compliance with legislative dictates. In particular, the AI Act classifies AI systems according to the risk they generate for users, establishing specific obligations for providers and users, which vary according to the level of risk identified. Similarly to Articles 25 and 35 of the GDPR, the AI Act requires providers to adopt a risk management system that maps and mitigates risks, informs stakeholders, and

[239] Voigt, P., & Hullen, N. (2024). Which Requirements Apply to GPAI Models?. In The EU AI Act: Answers to Frequently Asked Questions (pp. 129-149). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 138-149.

monitors the risk assessment system periodically. This enhancement of autonomous assessment by the provider responds to the logic of empowerment and accountability already present in the GDPR framework. However, given the extensive overlaps between the two disciplines, the simultaneous implementation of the GDPR and the IA Regulation could lead to overregulation. For this reason, it is essential that the relationship between data protection and artificial intelligence is carefully examined, in order to avoid conflicts or contradictions between the two regulations, which, if not properly harmonised, could result in conflict. In this regard, the European Commission, in the explanatory report accompanying the AI Regulation, clarified that the latter does not affect the applicability of the GDPR, which aims to protect a fundamental right, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union and Article 16 of the Treaty on the Functioning of the European Union. From this it follows that possible conflicts between the two regulations cannot be resolved by invoking the principle of speciality, since the GDPR protects a fundamental human right, provided for by a primary source. Therefore, both regulations will have to be applied and implemented jointly.

The AI Regulation and the GDPR have numerous points of contact in the way they are regulated. One significant similarity concerns the choice of legal instrument: the European Union has in fact adopted the form of a regulation for both regulations, rather than a directive, in order to ensure a uniform and binding regulatory framework for member states. Nonetheless, the points of contact between the two texts are not limited to the risk-based approach; in fact, the AI Regulation also recalls principles and mechanisms typical of the GDPR's data governance and management system, using them as tools to ensure correct and transparent training of AI systems. One area of considerable overlap between the GDPR and the AI Regulation concerns automated decision-making processes, which are governed by Article 22 of the GDPR, which regulates the processing of personal data carried out by means of automated decision-making processes. This provision represents one of the main points of contact with the AI Act, as it regulates processing such as profiling, which is frequently performed using artificial intelligence. It is important to note that while the AI Act assumes by default the adoption of automated decision-making processes - albeit with an appropriate level of human supervision - Article 22 of the GDPR establishes a fundamental right for data subjects not to be subjected to decisions that are based solely on automated processing, as these could have a significant impact on their legal sphere. However, Article 22 does not absolutely prohibit such processing, but provides an opening clause that allows it to be carried out under certain conditions, such as when it is necessary

for the performance of a contract, is authorised by EU law, or when there is the explicit consent of the data subject. An interesting aspect concerns the fact that although the GDPR establishes a relative prohibition for automated processing, allowing exceptions, it is the AI Act that explicitly prohibits certain automated decision-making processes. For instance, the AI Act prohibits social scoring, a practice that, according to the Privacy Rule, would be permitted provided adequate precautions are taken. This prohibition in the AI Regulation is more restrictive than the more permissive view of the GDPR, which, albeit with limitations, allows automated data processing in specific circumstances. The inconsistencies between the two regulations, therefore, raise important practical questions, especially in corporate contexts where business needs often drive the adoption of AI-based solutions. In such situations, the differences between the regulations could lead to difficulties in their joint application, resulting in potential conflicts between the compliance objectives set out in the GDPR and those outlined in the AI Act. Such tensions could be particularly relevant in contexts where the use of automated processes is essential for business efficiency, but must still comply with strict standards of data protection and data subjects' rights .[240]

In such a context, as analysed, *the Large Language Models*, in the economy of the EU Regulation 2024/1689 assume a particular relevance: it is, in fact, an element mentioned only once by the Regulation, in Annex XIII letter *d)*[241] , but which rises to paradigm with respect to the analysis and implementation of certain phenomena. In particular, the text of the law presents them as an example of one of the criteria to be considered in order to impute an IA system to the systemic risk referred to in Article 51 of the Regulation. But the *Large Language Models,*

[240] Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19, pp. 126-129.

[241] The text of Annex XIII, under the heading "Criteria for the designation of general purpose AI models with systemic risk referred to in Article 51" reads as follows: "In order to determine whether a general purpose AI model has equivalent capability or impact to those referred to in Article 51(1)(a), the Commission shall take into account the following criteria:

a) the number of parameters of the model
b) the quality or size of the data set, e.g. measured by tokens
c) the amount of computation used to train the model measured in floating-point operations or indicated by a combination of other variables such as the estimated cost of training, the estimated time required for training or the estimated energy consumption for training
d) the input and output modalities of the model, such as text-to-text (large language models), text-to-image, multi-modality and peak thresholds for determining high impact capabilities for each modality, as well as the specific type of input and output (e.g. biological sequences)
e) benchmarks and assessments of the model's capabilities, including taking into account the number of tasks that do not require additional training, the ability to learn new distinct tasks, the level of autonomy and scalability and the tools it has access to
f) whether the model has a high impact on the internal market in view of its reach, which shall be assumed when the model has been made available to at least 10 000 registered business users established in the Union
g) the number of registered end-users."

can also serve as a paradigm to better explain how the profiling activity and the issue of vulnerability already outlined (see *above*) represent also in the context of the Regulation under consideration a still open challenge for the jurist. The final text of the AI Act is the first law to mention the concept of human digital vulnerability numerous times and with a mature approach. Excluding those concerning cybersecurity, the AI Act makes 17 derivative references (e.g. vulnerability, vulnerable) to the original term. The references are in 9 recitals and 6 articles in the chapters of the regulation. Artificial intelligence systems that exploit human vulnerabilities on the basis of a person's age, disability or socio-economic circumstances (and produce significant harm) are prohibited. Human vulnerabilities are a parameter that the European Commission may consider when updating the list of 'high-risk artificial intelligence systems'. These vulnerabilities must be analysed and mitigated in the fundamental rights impact assessment, which is mandatory for suppliers of high-risk AI systems and some distributors. The 'codes of conduct' drawn up should assess and prevent the negative impact of AI systems on vulnerable persons. In the context of regulatory sandboxes, it is possible to process personal data beyond the original purpose in order to train and test artificial intelligence systems for the public interest, but only if (among other conditions) data subjects in a vulnerable condition due to their age or disability are 'adequately protected'. Despite all these references to human vulnerability in the AI Act, there is still considerable uncertainty about the concept. Section 3 of the AI Act contains 68 definitions, none of which relate to human vulnerability. Moreover, the language referring to vulnerability is varied throughout the text of the Act[242] . In particular, some doctrine[243] suggests that in

[242] Malgieri, G., & Rebrean, M. L. (2024). Vulnerability in the EU AI Act: building an interpretation. Available at SSRN 5058591. In particular, the authors identify six different formulations within the AI Act concerning vulnerability; the authors themselves organise these examples of vulnerability into five categories: (i) demographic characteristics, (ii) groups, (iii) personal sources of vulnerability, (iv) contexts and (v) no example. The first category, demographic characteristics, includes specific attributes that define and influence the life experience of each individual. These attributes may include factors such as age, gender, race, ethnicity and other personal traits that shape a person's experience. The second category, groups, relates to historical, protected and collective identities or affiliations that are formed through shared attributes or experiences between individuals. These groups may be linked to common characteristics such as culture, religion, language or geographical origin, which influence how people interact with the world and with each other. The third category, personal sources, includes traits that are both personal and volatile, that is, temporary and influenced by social, spatial, temporal, institutional, political, economic and environmental variables. These traits may vary over time and circumstances, such as a person's health status, work situation or social environment, and may influence an individual's vulnerability at certain times or contexts. Unlike the other categories, the fourth category, contexts, is not directly associated with personal or group attributes. Rather, it refers to situations in which different contexts are related to vulnerability, showing scenarios that either inherently determine vulnerability or are predisposed to generate it. For instance, an unstable economic or political environment may be a context that fosters vulnerability in people working in it. Finally, the category without examples is intended for those situations where no particular illustration of vulnerability is given, indicating that in some cases no concrete or specific examples have been defined to fully understand how vulnerability may manifest itself.

[243] See also Malgieri, G., & Rebrean, M. L. (2024). Vulnerability in the EU AI Act: building an interpretation. Available at SSRN 5058591.

order to explain the notion of vulnerability found in EU Regulation 2024/1689, it is possible to analyse three different implications that find their way into the legislative context under consideration. These are the vulnerable position, understood as a 'universal approach' to vulnerability; vulnerable groups, according to an approach based on identity perspectives; and vulnerability as a synthesis of the two previous terms. Dwelling on the concept of profiling, as defined by the GDPR (see *above*), and also considering it in the context of the Artificial Intelligence Regulation, one can look at the 'reasonably informed natural person', as referred to in Recital 132) and Art. 50(1): just like the consumer in relation to profiling, this individual, defined as an abstract figure, does not exist in practical reality, as online actions are generally influenced by limited rationality. In other words, the decisions made by individuals on the Internet are not always based on complete information or rational evaluations, but rather on cognitive processes that are influenced by limitations. In light of these analyses, it is difficult to draw a clear distinction between a natural person who can be considered 'reasonably informed', as stated in the Recital, and a natural person who is part of a vulnerable group[244] . That being said, without prejudice to the guarantees and protections envisaged by the Regulation in question, it seems clear that the *Large Language Models* can become a paradigm for a legal analysis that takes into account, on the one hand, the profiling referred to in Article 4(4) of the GDPR and, on the other hand, the intrinsic vulnerability of even the 'reasonably informed natural person'. Moreover, they lend themselves well to understanding the scope of the 'systemic risk' introduced by the AI Act. In general, it can be said that such models, while having great potential as they can be used in a large number of areas, still present a challenge to the jurist. The use of large language models (LLMs) for the protection of personal data can generate positive impacts, if these tools are designed, developed and deployed in a responsible manner, respecting the fundamental principles of data protection, privacy, human control and transparency. One area where LLM could prove useful concerns the detection of personal data within unstructured data, such as text contained in free fields. Although the manual detection of personal information in such contexts is relatively simple for humans, the automation of such a process through the use of simple rules is complex and not very scalable, especially when dealing with large files or data extracted from the Internet. In these cases, natural language processing, typical of LLMs,

[244] On this point, see more extensively, Malgieri, G., & Rebrean, M. L. (2024). Vulnerability in the EU AI Act: building an interpretation. Available at SSRN 5058591 ; Helberger, N., Sax, M., Strycharz, J., & Micklitz, H. W. (2022). Choice architectures in the digital economy: Towards a new understanding of digital vulnerability. Journal of Consumer Policy, 1-26.

could facilitate the identification and management of personal data, and also reduce the presence of such data in the training sets of LLMs themselves through automatic censoring or obfuscation techniques. However, the use of LLMs is not without risk and may lead to significant negative impacts on the protection of personal data. The formation of these models relies on huge amounts of data, which may include personal information extracted from public sources, such as the vast Common Crawl dataset, which gathers text from billions of web pages. Such data may concern not only public figures, but also private individuals, with the possibility of inaccurate or incorrect data being included. Managing such risks through appropriate protection measures is extremely complex, especially considering that, if not adequately protected, LLM could generate results that reveal sensitive information in the training data, with the consequent risk of privacy violations. Another problematic aspect concerns the so-called 'hallucinations' of models, i.e. the generation of erroneous information that, although appearing to be correct, may damage the public image of individuals and negatively influence decisions concerning them. Training on datasets containing biases or distortions may also lead to the perpetuation of discriminatory stereotypes, violating the principle of fair processing of personal data. Moreover, the implementation of data subjects' rights is particularly difficult in the context of LLM, as these models do not store data as in a traditional database, but store them through billions of parameters. Consequently, operations such as rectification, deletion or even the simple request for access to data learned from the model become problematic, if not impossible, especially since such data may be distorted or erroneous due to the 'hallucinations' generated by the model. Therefore, the adoption of LLM in the field of data protection requires careful thought and preventive measures to balance the benefits of automation with the potential risks related to the management and protection of personal information[245] . Indeed, those phenomena (see *above*) traditionally linked to the reading of Art. 4(4) in conjunction with Art. 22(1) of the EU Regulation 2016/679 such as *nudging, clustering, dark patterns*, although expressly forbidden by the Artificial Intelligence Regulation, could reappear under another, more subtle and ambiguous form. Transversely, the issue would all the more involve all individuals defined as *vulnerable*. Indeed, *Large Language Models* have been shown to excel at various tasks, including personalised recommendations. Existing assessment methods often focus on predicting assessments, relying on regression errors between actual and predicted assessments. Customisation tailors interactions, content or system recommendations to

[245] https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en .

individual users by analysing their behaviour, preferences and characteristics. It is crucial in areas such as content recommendation, user simulation, customised *chatbots*, user profiling, healthcare and education. *Large Language Models* excel in various natural language tasks, showing emerging capabilities. To align LLM results with individual user preferences, personalisation has become a key research objective, making benchmarks necessary for evaluation. Personalised recommendation, with its abundant data on user behaviour and preference signals, is adopted for the purpose of evaluating the personalisation of LLMs. The evaluation of personalised recommendation can typically be classified into evaluation-based and classification-based paradigms. In rating-based assessment, models predict a user's rating for an item and compute regression errors such as MAE and RMSE against actual ratings. However, user evaluation *bias* and query item quality are two influential factors on a user's evaluation scores for an item, which might prevent learning personal preferences in user-element pair data[246] . In general, it can be said that in LLMs, each user must have a *rating* history, i.e. an evaluation history, sufficient to be able to achieve and enable effective customisation. This also includes, for example, obtaining an *output* that contains an intrinsic text classification, obtained by peer comparison. Similarly, the *pairwise ranking paradigm* can be used to classify users on the basis of their preferences and on what they search for. As can be seen, as the AI employed becomes more refined, it becomes difficult to adhere to the prohibition laid down by the AI Act, as it is once again necessary to speculate on the meaning and limits of manipulation. In this context, it is therefore once again important to consider the issue of *bias*. It is possible to mitigate this phenomenon in the operation of *Large Language Models* and also in the judgement and evaluation phase inherent in them. This would lay the foundations for an effective implementation, in a completely technical and technological manner, of the Regulation under review. (Idem and see article in general). On this point, it should also be pointed out that the data fed to the AI can also be used for a *de-biasing* operation of the AI itself. This would make it possible to overcome the problems connected with profiling and personalisation directly by operating on the technology itself .[247]

5. Prospects *de iure condendo*: safeguards, *vexatae quaestiones* and opportunities: a 'systemic vulnerability'?

[246] On this point, see more extensively Tan, Z., Zeng, Q., Wu, Z., Liu, Z., Mo, F., & Jiang, M. (2025). Can Large Language Models Understand Preferences in Personalized Recommendation?. arXiv preprint arXiv:2501.13391.
[247] van Bekkum, M. (2025). Using sensitive data to de-bias AI systems: Article 10 (5) of the EU AI act. Computer Law & Security Review, 56, 106115, pp.2-4.

As for all regulatory instruments intended to provide a legal framework for complex and evolving landscapes, such as the one outlined by artificial intelligence, also the AI Act cannot escape de iure condendo reflections and assessments, which would favour a deeper internalisation and application. Indeed, the European Regulation poses a significant challenge in coordinating with existing regulations, especially in the area of personal data protection. On the one hand, the AI Act invokes, as is well known, principles of accountability that are reflected in the way personal data are processed, obliging those involved to demonstrate their compliance with legal provisions and to ensure adequate risk management. This approach, while guaranteeing a flexible and constantly evolving risk management system, contrasts with the GDPR, which, in its framework, imposes continuous adaptability and direct responsibility on the data controller to manage and update data processing processes. In other words, while the GDPR places an obligation of accountability and continuous adaptability on private entities, the AI Act delegates this responsibility more to the European legislator, who is charged with defining the AI systems to be considered high-risk and the relevant risk mitigation methods. This static nature of the AI Act's approach could be a criticism, especially considering the high speed of innovation and change typical of the artificial intelligence sector. A further element of criticism concerns the protection of the rights and freedoms of data subjects. Unlike the GDPR, which offers individuals a wide range of rights to be exercised vis-à-vis data controllers, with direct complaint tools to the supervisory authorities, the AI Act presents a decidedly narrower protection. In fact, the Regulation offers data subjects only two main remedies: on the one hand, the possibility to lodge a complaint with the market surveillance authority in case of violations, and on the other hand, the right to obtain explanations from manufacturers of high-risk AI systems regarding decisions taken through such systems, when these affect the rights and freedoms of individuals. The absence of a centrality of data subjects' rights, coupled with the lack of a clear liability regime within the AI Act, suggests that, in order to ensure adequate protection, the system of regulatory safeguards will necessarily have to integrate with the already existing data protection provisions, in particular those relating to the rights, reports and complaints under the GDPR. Critical issues also arise in cases where the AI Act refers to the GDPR. At these junctures, a more explicit clarification of the relationship between the two regulations would be useful to ensure better coordination and harmonious management between the disciplines. On the contrary, in cases where the Regulation does not explicitly refer to the GDPR, the impression might arise that some exemptions from the data processing law are permissible, which could create interpretative and applicative uncertainties. Finally, it should

be emphasised that the AI Act, although one of the many regulatory interventions of the European legislature on the digital single market, is part of a broader context of continuous evolution of digital law. In this respect, it would be desirable for the European legislator, in pursuing its regulatory coordination work, to implement contextual amendments not only to the text of the AI Act, but also to that of the GDPR, in order to improve the interrelation between the two legislations and simplify the overall system. An effective coordination strategy, aimed at codifying the numerous regulatory sources that govern the digital sector, could avoid the risk of regulatory hypertrophy, making the legal landscape clearer and more transparent.First, it should be pointed out that the AI Act's regulatory framework appears inevitably rigid. Although it is a laudable intention to exhaustively map each type of artificial intelligence system, classifying it according to its level of risk, such an approach runs the risk, in the short term, of being anachronistic and outdated. The rapid and relentless evolution of technology in the field of artificial intelligence entails the risk that the provisions of the Regulation will no longer be aligned with the new technologies that are constantly emerging. Although the Regulation provides for a periodic review of the classification of AI systems, the risk remains that, in an environment of constant change, the regulatory provisions become obsolete. In this sense, although the AI Act, like the GDPR, adopts a legal system based on risk management, it differs from the GDPR in that it does not apply the principle of accountability, which obliges the data controller to demonstrate its compliance with the regulatory provisions and to responsibly manage the risks associated with data processing. This accountability gap could hinder adequate protection in an ever-changing sector, where transparency and accountability are essential . [248]

Especially with reference to the *Large Language Models*, it is clear from what has been outlined that - far from the superfetation of regulations - it is possible to interpret the one already present in a new sense that is more attentive to the actual centrality of the human being and to his intrinsic vulnerability, while favouring the advantages of technological progress. This is how, therefore, the context of the *General-purpose Artificial Intelligence Systems* and *Large Language Models* can become a paradigm for a reading that leads to the creation of a new category, in point of interpretation, that looks at a 'systemic vulnerability'. In other words, exactly as is the case for the systemic risk envisaged by EU Regulation 2024/1689, which represents a peculiar type of risk associated with the high-impact functionalities of generative

[248] Falletta, P., & Marsano, A. (2024). Artificial intelligence and personal data protection: the relationship between the European Artificial Intelligence Regulation and GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19, pp. 133-134.

artificial intelligence models (GPAI), which exert a significant influence on the EU market, due to their scope or potential negative effects, both concrete and reasonably foreseeable, on public health, security, public order, fundamental rights or society as a whole, one could also think of an interpretation that relies on a 'systemic vulnerability', i.e. closely linked to the use of Artificial Intelligences that are harbingers of such a risk and provide, in this sense, the relevant protections. On this point, it is also necessary to provide a concrete overview of the implications that *systemic vulnerability* might have in the area of private law. Upon closer examination, the aforementioned category appears to be applicable in several areas. First of all, as far as personality rights are concerned, even in relation to the so-called *new technologies*, the investigation and the context of reference move beyond the traditional issues offered by the protection of privacy and confidentiality. In fact, new facets emerge in an increasingly automated reality, or traditional categories such as the right to personal identity are merged with ever new rights. Moreover, centuries-old foundations of private law, such as consent, the choice by the individual to create, modify or extinguish legal situations within the legal system must necessarily be the subject of reflection and inevitable upheavals. The same can be said of contractual matters and liability: the traditional domain of the individual leaves room for *something else,* represented by Artificial Intelligence and the unprecedented results it achieves. The 'systemic vulnerability' is thus presented as a condition to be ascribed, in general, to the physical person inserted in a technological and automated context. It has been pointed out, in fact, that vulnerability can take on different nuances and meanings, and thus have different meanings from a legal point of view (see *above*). Borrowing the term and meaning of the adjective 'systemic' from the European Union Regulation on Artificial Intelligence, it does not seem erroneous to assume that the condition of vulnerability in the legal sense also derives mainly from the use and modalities of AI.

Conclusions

This work has been a valuable opportunity to reflect on the implications of technological evolution and automation with respect to fundamental rights, from a jurisprudential point of view. As has been pointed out, the issues that the Doctrine had raised more than a lustre ago with reference to the profiling *pursuant to* Article 4(4) of EU Regulation 2016/679 read in conjunction with Article 22(1) of the same text of the law are now presented in renewed guise, even though in some respects they have now been transposed and prohibited by the legislator. Also with reference to this, dealing with the subject from the point of view of the *vulnerable individual* appears to be of particular importance at this historical moment and given the novelties in the legislative field. As analysed, the European Regulation on Artificial Intelligence stands as a *unicum* in the global context and the legislation should be credited with having been able to combine different instances in a *body of* effective rules, even while awaiting the full entry into force of the Regulation. However, it should be noted that technological evolution proceeds at a very rapid pace and that it is difficult for the legislator, the interpreter, and the jurist to match this speed. From this point of view, the AI Act therefore tries to be very general in scope, but this can be a double-edged sword. In the economy of the paper, the focus has been on *Large Language Models,* since they represented - in the very last stages of the legislative *process* that led to the text of the Regulation just mentioned - a novelty with a disruptive scope. It was a privileged vantage point to be able to study, analyse and experiment them from an interdisciplinary point of view right from the very beginning. Similarly, the intersection with the profiles referred to in the GDPR and the analysis of the new vulnerabilities, has been part of a study that began more than a decade ago and that has had the privilege of looking at the phenomena from many facets and above all of following their evolution over time, grasping the criticalities, strengths, and points of development for the formulation of solutions and innovative interpretations. From the outset*, the leitmotif* of this study has been the focus on fundamental and personality rights, which must never be limited to formulas but must always be filled with meaningful content. A content that, on closer inspection, changes as the context and *res of* reference change; a content that, however, crystallised, can stand as a parameter of abstoric evaluation; a content that stands in the *mare magnum* of the evolution of law, and in this case of private law. It is of great interest to note how many of the institutes of this discipline have been renewed by the

great thrusts of recent decades, and it is interesting to note how private law itself is to be considered in a different perspective also from the geographical point of view. With respect to this last point, it is evident that - given the history and peculiarities of each system - private law now takes on the dimensions of supranationality. A supranationality that is coloured by the impulses and legal traditions of its constituent elements, also from a comparative perspective. On this point, it is worth mentioning how much the present work has benefited, even in this case, from research that is the fruit of comparison with doctrinal readings, interpretations, supra- and international jurisprudence. In this case too, the reflections that have emerged are more than five years old: one has reason to believe, in fact, that not only the institutes, but also the legal language - so different in all legal systems from the non-sectoral one and so reflective of *Civil Law* and *Common Law* traditions - has emerged renewed and, albeit partially, harmonised. This leads one to consider how, much more than in the past, the global interconnectedness is tangible and, this consideration allows one to analyse the first point put forward, concerning the renewal of private law. On this point, one can reflect on the epoch-making changes that have taken place even in the last five years and that have become the harbingers of dynamic and interconnected legislation.

1. The delicate and necessary balance between technological innovation and environmental protection concerns.

The issue of climate change has affected, among other things, the way to regulate new technologies and automation. The Regulation recently examined also refers to environmental concerns. It is, in fact, of great interest to note how technological development can, on the one hand, constitute a bulwark for environmental protection—through the implementation of human dynamics with increasingly limited time resources; but on the other, this same technological development can pose a threat to the environment through the massive use of natural resources and the exploitation of already deprived and overused global areas and populations. Just think, in this regard, of how many resources a large-scale linguistic model can employ for its operation.

In other words, the issue of climate change is increasingly impacting the ways in which emerging technologies and advanced automation systems are regulated. The most recent Regulation examined—placed within the framework of the European governance of algorithmic innovation and automation—contains explicit references to environmental sustainability requirements, a sign of the progressive integration of the principle of

sustainable development into digital infrastructure policies.

From an analytical perspective, technological development demonstrates a dual structural value. First, it can serve as a fundamental driver for environmental protection through the implementation of intelligent systems capable of optimizing complex production processes, reducing energy inefficiencies, implementing predictive models for natural resource management, and supporting public decisions based on high-resolution temporal and spatial data. Technologies such as machine learning models, advanced sensor systems, and industrial automation networks (IoT and IIoT) enable the transformation of time- and resource-intensive human activities into highly efficient operations, with potential mitigation effects on emissions and the ecological footprint of production sectors.

However, alongside this functional dimension, technological progress presents, as mentioned, a critical counterpart. Digital technologies characterized by high computational power require considerable volumes of material inputs—including electricity, water for cooling infrastructure, and critical raw materials for the production of semiconductors and dedicated hardware. The resulting environmental pressure is evident throughout the technology's entire life cycle, from extraction to decommissioning, including the construction of data centers and the daily operation of computing platforms. This impact is further amplified when resource extraction and infrastructure siting occur in contexts characterized by high socio-environmental vulnerability, with the risk of exacerbating the exploitation of populations and territories already exposed to degradation and marginalization.

A paradigmatic example is represented by Large Language Models, whose training phase requires an extremely high number of computational operations and, consequently, significant energy requirements. Even in the inference phase, especially in highly scalable contexts, these models continue to generate a significant energy burden. Their apparent immateriality conceals, in fact, an energy-intensive physical infrastructure that contributes to global resource consumption and emissions, necessitating a thorough technical and regulatory assessment based on various indicators.

In light of these considerations, regulation that integrates environmental sustainability criteria into the design, implementation, auditing, and monitoring phases of advanced technologies appears essential. This entails the adoption of technical standards, transparency requirements on energy performance, environmental impact assessment obligations, and accountability principles extended across the entire digital value chain. Only a systemic regulatory approach can ensure that technological innovation does not result in an expansion of the global ecological footprint, but rather contributes to a digital and AI transition consistent with

climate objectives.

2. Private law as a privileged field of analysis and evolution.

A further renewal, with reference to private law, can also be considered the progressive distancing from its "patrimonial" aspect, arriving at an increasingly depatrimonialized conception in which, among other things, issues regarding, for example, vulnerability can find ample scope for research and development. This evolution entails a downsizing of traditional private law categories in favor of a growing focus on the individual, their dignity, and the multiple forms of fragility that characterize contemporary life. From this perspective, issues such as individual and collective vulnerability, the protection of vulnerable individuals, the consideration of information and power asymmetries, as well as the impact of social and technological contexts on fundamental rights, find ample scope for research. The traditional vision of private law as a system of rules aimed at regulating exchanges is further enriched by an ethical-social dimension that aims to intercept new, not immediately identifiable, needs centered on the protection of the individual in his or her entirety.

Considering the evolution of technology and automation, new ways of interpreting reality emerge: consider, in this regard, issues such as so-called mental privacy and neurorights. These raise radical questions about the protection of thought, the cognitive integrity of the individual, and the possibility that brain-computer interface technologies impact the freedom of self-determination and personal identity itself. The law thus finds itself defining boundaries and guarantees in previously unexplored territory.

The institution of responsibility takes on particular importance, strongly challenged by the advent of Artificial Intelligence. This responsibility is being scaled back in its traditional structure and increasingly becoming a tool to be used for preventive purposes. Consider also the issues of consent and self-determination, which were first challenged by the major legal movement regarding the standardization of the digital landscape and automation. Consider how new media and art—and the related rights—are interconnected with Artificial Intelligence. All this illustrates how academic and legal research continues to be a key and necessary tool for grasping the precipitate of reality and for approaching it—and the related regulatory, jurisprudential, and practical innovations—in a vigilant, critical, and creative manner.

The questions recently considered remain partly unanswered even in the latest Italian legislation on the subject, Law No. 132/2025.

Ultimately, every phenomenon can become a research paradigm and, above all, raise questions, issues, and essential starting points for the advancement of knowledge. Therefore, this research project, after answering questions, offering solutions, and analyzing issues, intends to present itself as a starting point and a question.

BIBLIOGRAPHY AND OTHER SOURCES CONSULTED

Antonella Tartaglia Polcini, Relazione introduttiva CONVEGNO REALTÀ PSICHICA, REALTÀ DIGITALE: UN'INTEGRAZIONE POSSIBILE? - 23 febbraio 2024 Benevento

Bishop, C. M., & Bishop, H. (2023). Deep learning: Foundations and concepts. Springer Nature.

Bravo, F. (2021). Intermediazione di dati personali e servizi di data sharing dal GDPR al Data Governance Act. Contratto e impresa Europa, 1(1), 199-256.

Bravo, F. (2022). «Destinatario» dell'informazione e trattamento dei dati personali nell'evoluzione dell'ordinamento europeo. In M. D'Auria (a cura di), I problemi dell'informazione nel diritto civile, oggi. Studi in onore di Vincenzo Cuffaro (pp. 431-454). Roma Tre-Press

Busia, G. (2013). Le frontiere della privacy in Internet: La nuova corsa all'oro per i dati personali. Internet: Regole e tutela dei diritti fondamentali.

Bygrave, L. A. (2001). Automated profiling: Minding the machine: Article 15 of the EC data protection directive and automated profiling. Computer Law & Security Review, 17(1), 17-24.

Cavoukian, A., Dix, A., & El Emam, K. (2014). The unintended consequences of privacy paternalism. Information and Privacy Commissioner of Ontario, Canada.

Ciullo, M. Profiling and automated decisions: theoretical and evolutionary profiles, in corso di pubblicazione.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., ... & Helbing, D. (2012). Manifesto of computational social science. The European Physical Journal Special Topics, 214, 325-346.

Corsi, R., & D'Albergo, E. (2024). La politica dell'intelligenza artificiale general purpose: immaginari sociotecnici, democrazia e policy frame nel processo decisionale della regolazione europea ("EU AI ACT"-2022-2024). IM@GO, 23, 109-130.

D'Ippolito, G. (2021). Profilazione e pubblicità targettizzata on line: real-time bidding e behavioural advertising. Edizioni scientifiche italiane.

Daelman, C., & Yordanova, K. (2023). AI through a human rights lens. The role of human rights in fulfilling AI's potential. Artificial intelligence and the law. Intersentia, Cambridge, 135-168.

De Vivo, I. (2024). Il potere d'opinione delle piattaforme-online: quale ruolo del "regulatory turn" europeo nell'oligopolio informativo digitale?. Federalismi.it, 2, 45-75, p.263.

Ebers, M. (2019). Regulating AI and robotics: ethical and legal challenges.

Fabiano, L. (2024). Algorithmic Recommendation ed Artificial Intelligence nell'esperienza statunitense ed europea fra esigenze strategiche e tutela dei diritti fondamentali. IRCOCERVO, 23(1), 440-456.

Falletta, P., & Marsano, A. (2024). Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR. Rivista italiana di informatica e diritto, 6(1), 19-19.

Flew, T. (2019). The platformized internet: Issues for media law and policy. Disponibile su SSRN 3395901.

Formici, G. (2021). La disciplina della data retention tra esigenze securitarie e tutela dei diritti fondamentali: Un'analisi comparata (Vol. 94). Giappichelli.

Frosini, T. E. (2020). Il costituzionalismo nella società tecnologica. Giurcost.org, Liber Amicorum per Pasquale Costanzo, 25.

Helberger, N., Sax, M., Strycharz, J., & Micklitz, H. W. (2022). Choice architectures in the digital economy: Towards a new understanding of digital vulnerability. Journal of Consumer Policy, 1-26.

Jo, T. (2021). Machine learning foundations. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-65900-4

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big data & society, 1(1), 2053951714528481.

Malgieri, G. (2023). Vulnerability and data protection law. Oxford University Press.

Malgieri, G., & Rebrean, M. L. (2024). Vulnerability in the EU AI Act: building an interpretation. Disponibile su SSRN 5058591

Manheim, K., & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. Yale JL & Tech., 21, 106.

McStay, A. (2018). Emotional AI: The rise of empathic media.

Messinetti, R. (2021). Comunicare nell'infosfera. La vulnerabilità della persona digitale. FEDERALISMI.IT, (18 del 28/07/2021).

Montuori, L., & Siano, M. (2017). Evoluzione del concetto di consenso informato nel mondo digitale e transizione del marketing tradizionale alle attuali sfide della profilazione. Le nuove frontiere della privacy nelle tecnologie digitali: bilanci e prospettive, 101-126.

Natale, G. (2024). Il nuovo regolamento europeo AI Act. Diritto di Internet: Intelligenza artificiale, digital copyright e data protection, 2, 123-145.

Novelli, C. (2024). L'Artificial Intelligence Act Europeo: alcune questioni di implementazione. FEDERALISMI.IT, 2, 94-113.

OECD (2019). Artificial Intelligence in Society. OECD Publishing, 11 giugno 2019.

Pelino, E. (2023). L'interazione tra DSA e GDPR. In L. Bolognini, et al. (a cura di), Digital services act e Digital markets act: definizioni e prime applicazioni dei nuovi regolamenti europei. Giuffrè.

Perlingieri, P. Relazione conclusiva SCUOLA ESTIVA La persona umana nell'epoca della "doppia transizione digitale ed ecologica" – San Benedetto del Tronto (AP), Settembre 2024.

Pierucci, A. (2019). Elaborazione dei dati e profilazione delle persone. I dati personali nel diritto europeo.

Pizzetti, F. (2016). Privacy e il diritto europeo alla protezione dei dati personali: Dalla Direttiva 95/46 al nuovo Regolamente europeo. G Giappichelli Editore.

Poletti, D. (2018). Regolare la tecnologia: il Reg. UE 2016/679 e la protezione dei dati personali. Un dialogo fra Italia e Spagna. STUDI IN TEMA DI INTERNET ECOSYSTEM, 1-512.

Purtova, N. (2014). Default entitlements in personal data in the proposed Regulation: Informational self-determination off the table… and back on again?. Computer Law & Security Review, 30(1), 6-24.

Rodotà, S. (1998). Relazione per l'anno 1997 del garante per la protezione dei dati personali. Informazione e informatica.

Rossi Dal Pozzo, F. (2018). La tutela dei dati personali nella giurisprudenza della Corte di giustizia. EUROJUS.

Sah, S. (2020). Machine learning: a review of learning types.

Sarti, O. S. (2017). Profilazione e trattamento dei dati personali. Innovazione tecnologica e valore della persona-Il diritto alla protezione dei dati personali nel regolamento UE 2016, 679, 573.

Schulze, R. Digital Vulnerability in European Private Law – Conclusions, in corso di pubblicazione.

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.

Sola, A. (2023). Ambiti di interesse per la regolazione delle economie dei dati nel rapporto tra diritto e tecnologia. FEDERALISMI.IT, (10), 195-217.

Solove, D. J. (1880). Privacy Self-Management and the Consent Dilemma (2013). Harvard Law Review, 126, 1880.

Sposini, L. (2024). La Democrazia nell'Era delle Piattaforme Digitali tra Profilazione e Personalizzazione dei contenuti. Il Paradosso dell'Inconsapevolezza.

Sunstein, C. R. (2018). Republic: Divided democracy in the age of social media.

Tan, Z., Zeng, Z., Zeng, Q., Wu, Z., Liu, Z., Mo, F., & Jiang, M. (2025). Can Large Language Models Understand Preferences in Personalized Recommendation? arXiv preprint arXiv:2501.13391.

Van Bekkum, M. (2025). Using sensitive data to de-bias AI systems: Article 10 (5) of the EU AI act. Computer Law & Security Review, 56, 106115.

Voigt, P., & Hullen, N. (2024). Which Requirements Apply to GPAI Models?. In The EU AI Act: Answers to Frequently Asked Questions (pp. 129-149). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wong, B. (2021). Online personalised pricing as prohibited automated decision-making under Article 22 GDPR: a sceptical view. Information & Communications Technology Law, 30(2), 193-207.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

Zhou, Z. H. (2021). Machine learning. Springer Nature.

GPDP, 'Fidelity cards' and consumer guarantees. The Garante's rules for fidelity programmes, 24 February 2005, web doc. no. 1103045

GPDP, Guidelines on the processing of personal data for online profiling, 19 March 2015, web doc. no. 3881513.

CGUE, C-582/14, 19/10/2016

1 GPDP; Order No. 488 of 24 November 2016. Web platform for the development of reputational profiles, web doc. no. 579683.

CGUE, C-434/16, 20/12/2017

Court of Cassation, sec. I Civ. - 25/05/2021, n. 14381

CGUE, C-184/20, 01/08/2022

LEGISLATION AND OTHER MEASURES

Annex III, point 5(b) AI ACT

Annex III, point 5(c) TO ACT

Art 243 GDPR

Art 25 GDPR

Art 40 GDPR

Art 41 GDPR

Art 43 GDPR

Art. 127 EU Regulation 2016/679

Art. 13 AI ACT.

Art. 14 Legislative Decree 196/2003

Art. 15 GDPR 1995/46/EC

Art. 17 GDPR

Art. 2 AI ACT

Art. 2 EU Regulation 2016/679

Art. 2055 of the Civil Code

Art. 22 GDPR.

Art. 24 GDPR

Art. 28 DSA

Art. 3 AI ACT

Art. 38 DSA

Art. 42 GDPR

Art. 46 par. 1 AI ACT

Art. 5 GDPR

Art. 5 EU Regulation 2024/1689

Art. 53 AI Act

Art. 8 European Convention on Human Rights

Art. 8 GDPR

Art. 82 GDPR

Art. 97 EU Regulation 2024/1689.

Art.26 DSA

Art.6 AI ACT

Articles 9 and 10 EU Regulation 2016/679.

Article 10 ECHR, Article 19 UDHR and Article 19 ICCPR.

Article 11 ECHR, Article 20 UDHR and Articles 21 and 22 ICCPR.

Article 16 Treaty on the Functioning of the European Union (TFEU).

Article 2 European Convention on Human Rights (ECHR)

Article 27 of Directive (EU) 2016/680

Article 3 Universal Declaration of Human Rights (ECHR).

Article 3 UDHR

Article 35 of Regulation (EU) 2016/679

Article 5 ECHR

Article 6 ECHR, Articles 6, 7, 8 and 10 UDHR and Article 14 ICCPR

Article 9 ECHR, Article 18 UDHR and Article 18 ICCPR.

Article 9 ICCPR.

CHAPTER III AI ACT

Recital 10) AI ACT

Recital 110) AI ACT

Recital 115) AI ACT

Recital 118) AI ACT

Recital 27) EU Regulation 2016/679

Recital 30) GDPR

Recital 42) AI ACT

Recital 48) AI Act

Recital 53) AI ACT

Recital 53) AI ACT

Recital 59) AI ACT

Recital 75) GDPR

Recital 9) of the GDPR

Convention 108 for the Protection of Personal Data of 1981

Article 7 of EU Regulation 2016/679

Directive (EU) 770/2019

Directive (EU) 771/2019

Directive 95/ 46/EC of the European Parliament and of the Council. Of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (so-called "parent directive").

Article 4 GDPR

Article 3(2)(b) of Convention 108 for the Protection of Personal Data of 1981

Preamble of the Universal Declaration of Human Rights

Regulation (EU) 2016/(,79 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation).

Regulation (EU) 2022/1925

Regulation (EU) 2022/2065

Regulation (EU) 2023/2854

SITOGRAPHY

https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_it

https://digital-strategy.ec.europa.eu/it/policies/regulatory-framework-ai

https://eur-lex.europa.eu/IT/legal-content/summary/the-eu-s-new-digital-single-market-strategy.html

https://www.agendadigitale.eu/industry-4-0/ia-limpatto-dei-large-language-model-sulle-aziende-usi-e-problemi/

https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en