ENGLISH SUMMARY[1] OF THE BOOK TITLED

ARTIFICIAL INTELLIGENCE AND THE SEMANTICS OF CHANGE: A CRITICAL READING©

(ORIGINAL TITLE IN ITALIAN: *INTELLIGENZA ARTIFICIALE E SEMANTICA DEL CAMBIAMENTO: UNA LETTURA CRITICA*)

AUTHOR OF THE BOOK: SILVIA SALARDI[2]

PUBLISHER: GIAPPICHELLI

YEAR: 2023

PAGES: 152

Chapter 1

Technological progress and semantic issues: *diffinitio* and *partecipatio*

1. Introduction

Artificial Intelligence (AI), together with the environmental question, is today one of the most debated themes across disciplines and one that receives the greatest attention from national, European, and international institutions. What is at stake is technological *sovereignty*. Whoever first gains control over computer structures and over the regulation of intelligent technologies will obtain decisive advantages in the economy, in directing human activities, and in shaping a particular model of society. The great world powers are competing in this race.

In Europe, this competition has been taken very seriously, especially in defining rules aimed at preventing the uncontrolled introduction of AI systems into the market, with the risk of violating fundamental rights and freedoms, and at indirectly compelling other global competitors to adopt similar regulations (extraterritorial reach). With this double objective, the European Union elaborated a Proposal for a Regulation on AI, an important step that we will occupy ourselves with in this work. This regulatory achievement was reached in a context where the dialectic between the human and the machinic is still fed by threatening oppositions in both media and institutions. Its legal trace is recent: in 2017, the European Parliament, caught in this same dialectic, proposed granting robots legal personality (e-personality).

This dialectic, which we call the *dialectic of opposites*, sets the human being, whose destiny seems directed toward marginality, against the machine, imagined as capable of replacing the human in nearly all activities. At its base lies an anthropomorphizing ideology of the machinic, an

[1] English translation and summary by Douglas Luis Binda Filho, PhD candidate University of Milano-Bicocca. This translation is intended to give an overview of the main aspects discussed in the book and has not to be considered exhaustive of all the ethical and legal knots present in the book. All rights are reserved©.

[2] Silvia Salardi is associate professor in Philosophy of law and Bioethics at the School of Law of the University of Milano-Bicocca. For further information on the author, please visit the following links:
https://philet.giurisprudenza.unimib.it/
https://en.unimib.it/silvia-salardi

interpretative process *per relationem* that reads all forms of reality by analogy with human behavior. This applies to ancient interpretations of God and Nature and today to AI.

This mode of explanation contains two errors. The methodological one consists in taking the human brain, rational and emotional, as a model for constructing intelligent machines, committing a naturalistic fallacy when this analogy is extended to the value level. The epistemological one lies in exalting machines while diminishing human potential, obscuring the real limits and possibilities of both, and allowing a reductionist ideology to persist.

This dialectic, long present in literature, cinema, and common sense, and reflected in European institutions, must be examined to seek its reconciliation or overcoming. One of its main sources is language. The ubiquity, vagueness, and ambiguity of the notions and terms used in debates sustain this dialectic. Hence the need to cleanse language of the anthropomorphizing and reductionist ideology that has accompanied AI since its beginnings. From literature and cinema this ideology has spread through language into common sense, influencing knowledge and institutions.

Ambiguous and vague concepts can become linguistic traps that keep the dialectic of opposites alive as a covert form of power, compromising rational public choices and obscuring fair evaluation of AI's advantages, limits, and benefits. Recognizing the malleability of language, especially ordinary language, and the historical layering of meanings, this work proposes to reflect on the current use of key notions in the public and institutional narrative on AI and on how such uses shape policies and visions of the machinic and the human.

As Uberto Scarpelli wrote, the aim is to "restore language as a good means of expression, communication and orientation, good not according to criteria internal to language, but good for man with his current attitudes, needs and projects in the circumstances in which he now finds himself". The focus on language is justified for two reasons. It is the main means of human communication and influence, and the same technologies discussed here are built on linguistic structures. This is evident in Human Language Technologies, also called language technologies, whose use has raised ethical and legal problems such as linguistic underrepresentation, the digital extinction of underresourced languages, and the persistence of linguistic barriers. These issues show that the use of new technologies requires careful reflection on language to avoid discriminatory homogenization contrary to Article 22 of the Charter of Fundamental Rights of the EU. The field of linguistic analysis can extend to the concepts used in public and institutional narratives on AI, and this proposal of in-depth analysis aims to investigate two order of events connected to language and its social force, the transformations of human beings and the transformations of human beings'circumstances. Decades of this dialectic of opposites have deeply influenced not only the way people view the machinic but mostly the often reductionist idea they have formed of themselves and of the human species.

The analysis of the concepts used in the public narrative on AI can help restore narrative balance and open space for social and cultural visions alternative to the engulfing proposals of those who hold technological and economic power.

2. The advent of Artificial Intelligence and the semantics of change

Much has been written, and continues to be written, about the capacity of AI to revolutionize our way of life. Klaus Schwab speaks of a *fourth revolution*, referring to a transformation characterized by intelligent technologies capable of combining the physical, digital, and biological

spheres, and that, by their nature, will challenge the very meaning of what is human and of human nature. It is assumed that through AI it will be possible to bring about a radical change in our lives and in our nature itself. This assumption, which underlies the entire narrative of AI, but which does not necessarily imply the truth of its premise, that its intervention in almost all human activities will bring universally distributed well-being, is conveyed through deliberate semantic choices. Among these, two terms stand out: revolutionand innovation. They constitute the basis of the semantics of change that pervades our century, though they already took shape in the previous one. This section examines their semantic nuances and the visions of social organization they express.

Scientific and technological progress has shown over time two fundamental features: continuity, understood as constancy despite pauses, and irreversibility, which together attest to its reality. Innovation in science and technology unfolds as a continuous progression of knowledge, whose efficacy can be verified by its impact on people's quality of life. But not every step in this progression can be called revolutionary. Many great transformations of the past have simply confirmed or expanded earlier theses and ideas. A revolution, in the proper sense, takes place when the epistemic paradigm itself changes, as in the so-called scientific revolution. In other words, for a discovery or innovation to be revolutionary, there must be a *quid pluris*, a substantial disruption of our ways of thinking, interpreting, and representing reality. The scientific revolution, for example, replaced the traditional teleological view of nature with an empirical and causal one, forcing a redefinition of the human-nature relationship. Unlike what happens with AI, the major revolutions of the past were not defined as such by their protagonists. Figures such as Copernicus and Newton, though revolutionary, did not perceive themselves as such, not only because the term was not yet used in scientific language, but also because they regarded themselves as "heirs and rediscoverers of Antiquity". What later generations called revolutionary was, for them, an expansion of ideas already conceived by their predecessors.

This reminds us that the term revolution can be understood in two senses. In its full sense, it refers to a profound change of conceptual, cultural or social paradigm, disconnected from what came before. In a narrower sense, it may instead indicate new discoveries that simply extend or reinforce existing ideas. True scientific and technological revolutions are therefore rare. What we often call revolution is instead a necessary passage within a broader process of improvement: an innovation, significant but not a radical rupture. Clarifying this distinction is the first step toward conceptual precision. The second concerns the ambiguity produced by the interchangeable use of revolution and innovation in common language. These terms are often used as synonyms: technological revolution and technological innovation, digital revolution and digital innovation, algorithmic revolution and algorithmic innovation. This equivalence is problematic, since in common understanding revolution evokes an epochal upheaval, while its narrower meaning, as mere expansion of pre-existing ideas, tends to disappear. The result is a linguistic mechanism that amplifies the perception of the radical nature of the social transformations in progress and conditions public attitudes toward them.

When digitalization and the use of AI in human activities are presented as revolutionary, these processes appear to the public as both uncontrollable and inevitably positive. The identification between revolution and innovation conceals the fact that such transformations (the advent of AI and the digital reorganization of society) are neither spontaneous nor unexpected, but for the most part planned and governable.

In this restricted sense, revolution can indeed be identified with innovation, characterized by continuity and irreversibility, and therefore by the possibility of control and direction. For its part, innovation may also be understood in a broad or in a narrow sense. In ordinary language, it

designates "a modification introducing elements of novelty" aimed at improving people's living conditions. Such modification may concern very different sectors and contexts and involve diverse forces and instruments depending on the transformations one seeks to achieve. This general definition shows that innovation is not only technological: one can also speak of cultural, artistic, or social innovation.

In the current historical moment, however, the prevailing tendency in European institutional discourse is to identify innovation exclusively with technological innovation, leaving in the background the broader meaning of the term, which would include all forms of innovation oriented toward the improvement of human well-being.

In this semantic flattening of innovation onto technological innovation, and in the qualification of the latter as revolutionary, one can discern the implicit message of a future society modeled predominantly (if not entirely) on technocentric foundations. If innovation is assumed to be only technological and this innovation is deemed revolutionary, then technology appears to evolve autonomously, independent of social and cultural events, and thus capable of progressing through radical ruptures with the past. This understanding of technological innovation is accompanied by an optimistic vision that connects technological progress to changes presumed to be always positive for human existence, that is, for all humans. From this assumption follow practical choices concerning investment in the training of those who will inhabit the technologically renewed society. Economic resources are therefore directed mainly, if not exclusively, toward creative and productive activities directly linked to technological development, that is, to technological progress believed to revolutionize human life for the better.

Although language is not the only vehicle through which social and cultural models are conveyed, it is true that in the narrative of AI the constant overlap between innovation and revolution, and the reduction of innovation to its technological meaning, promote an attitude of particular favor toward technical-scientific disciplines, the main actors of these transformations. The risk of such an emphasis is to overlook that governing technological progress requires abilities and critical capacities that stem from other fields of knowledge, notably the humanistic ones.

The problem is not negligible, since the restriction of innovation to its technological meaning has already produced political choices, for instance, in education policy, aimed at cutting costs in areas that do not train future technologists or technicians. The critical concern does not lie in the preparation of such figures, but in the increasingly explicit tendency, long denounced, to marginalize disciplinary perspectives essential for developing critical thinking. Critical thinking allows for the right distance from present events and for a long-term vision of choices often dictated instead by immediate economic or profit motives.

Summing up these reflections, the interchangeability between innovation, understood only as technological progress, and revolution in its most radical sense, especially when sanctioned by institutional and political language, leads the recipients of this discourse to adopt, often unconsciously, a favorable and selective stance toward a phenomenon (the development of AI) whose long-term effects on society and individuals remain uncertain. Clarifying, instead, that today's digital transformation is not a revolution in the proper sense but a profound innovation that originates in technological advances and has been largely planned over time, makes it possible to expose the contrasts and tensions at the basis of the dialectic of opposites, which is nothing other than the confrontation between ethical and value-laden visions. If one highlights that the definitional and semantic issues surrounding the terms used in the AI debate are not empty terminological disputes, it becomes possible to reveal the diversity of ends that technological

knowledge, nourished by other forms of knowledge, can pursue. Taking the position that today's technological progress is innovative but not revolutionary in the proper sense allows us to recognize that it is built on a long-term path of choices and explorations of possibilities, including certain cultural, social, ethical, and value perspectives while excluding others.

In this innovative trajectory, to which technology is now given an exclusive role, there is the risk of leaving aside factors, methods, and cultural contributions that many advocates of equality and social justice have long sought to include in their proposals accompanying technological progress.

We can therefore conclude that the use of the term revolution in the narrative of AI is instrumental to a certain way of framing technological development and of realizing specific ethical visions of social coexistence. It refers to a vision of society in which efficiency and profit prevail above all else. The transformation of industrial society into an information society aims precisely at these goals. In this model, "computers manage the storage, organization, and retrieval of information, govern every kind of machine, control work flows, augment reality with virtual objects and contents, animate physical and virtual automata". For this reason, one speaks of "a new socio-technological paradigm" and of a fourth revolution, which concerns not so much a fourth industrial revolution driven by AI as a radical modification of the human being centered on the infosphere. However, this revolution should be further specified to avoid the linguistic traps linked to its ordinary use. It has not overturned the world. It lacks that *quid pluris* which would make it a revolution in the proper sense. It appears, rather, as the fulfillment of a long-planned process. Emphasizing that it is not an unforeseen event that has overwhelmed humanity but a development long in motion dispels the idea of a phenomenon governed by external forces before which individuals are powerless. Pointing out that this so-called revolution is not a rupture but an ongoing innovative process also reveals a cognitive bias that often affects human reasoning, limiting our capacity to judge and discern, the framing effect, whereby the same decision problem, if presented differently, leads to different choices. The decisive factor is not the informational content but the way it is framed. This bias is closely linked to language and to the tendency to frame messages in emotionally charged terms. When the transformation brought by AI and information technologies is presented as a revolution, as rupture rather than as a planned and guided process, it becomes easier to convey the idea that what is happening is inevitable, a destiny beyond regulation or limits. Yet the phenomenon we are witnessing has historical roots, follows a path traced by human decisions, and can only be understood once we abandon the notion of a neutral technology that arises and evolves solely for the good of humankind. Technological development, both past and present, together with its philosophical and political implications, must be situated within the economic and commercial mechanisms governing social relations. As Shoshana Zuboff observes, "we cannot assess the course taken by the information civilization without understanding that technology is not and cannot be a thing apart, isolated from economy and society". Even if the current phenomenon is to be regarded as an innovative process and not a revolution in the radical sense, this does not exclude the need to adapt through interpretation or through new elaboration the conceptual categories, particularly in the legal field, required to confront the unprecedented challenges brought by AI. Zuboff notes, for instance, that the triumph of surveillance capitalism lies precisely in its unprecedented nature: when confronted with something without precedent, we interpret it using familiar categories and thereby render invisible its novel features. To orient the advent of AI in directions that unmask the *ideology of inevitability*, it is indispensable, both semantically and conceptually, to verify whether the ethical and legal categories at our disposal remain adequate, and, where they are not, to reconstruct or elaborate them anew so as to "adjust and improve language, making it once again a good instrument for human beings, responsive to their present attitudes, needs, and projects".

An additional reason that can be adduced in favor of interpreting the digital and computerized transformation of society as the completion of a largely planned path, rather than as a rupture with the past, lies in its connection with a philosophical tendency within the Western cultural tradition, an unrelenting tendency to eliminate uncertainty and ensure control. It is a tendency that feeds and reinforces itself through an objectivist rhetoric that defines reality, including human behavior, only in terms of intelligible data structures, self-evident and reducible to a clear and simple explanation. Taking this tendency into account, which Dreyfus identifies as an *ontological assumption*, one can hypothesize that excessive emphasis on the revolutionary scope of AI, understood in its most drastic and dramatic sense, may appear ideological, as it conceals from the recipients of this transformation its planned and value-based origin.

One of the main difficulties encountered by the public addressed by AI's narrative, shaped by the dialectic of opposites, lies in its inability to perceive the reductionist and objectivist visions of the human underlying that dialectical framework. This can have serious repercussions on both individual and collective attitudes. These pervasive visions, deeply rooted in culture, often contribute to indifference and resignation in those who feel subject to fate, discouraging the motivation and the reasons to maintain a proactive stance toward life events. The dialectic of opposites, nourished by preconceived anthropomorphism, plays a decisive role in the tendency toward a "univocal and definitive schematization" of human experience, leaving no room for the language of freedom or for an ethics of responsibility.

Speaking of innovation as revolution, implying the drastic sense of the term, or clarifying instead that the advertised revolution is merely the culmination of a long process, profoundly changes the analytical and perceptive perspective on current phenomena. In the first case, the path seems already traced. In the second, it appears traceable through the contribution of rules capable of balancing the different interests at stake. Put differently, if technological progress is equated with a sudden and overwhelming event such as a revolution, the implicit message is that nothing can be done in advance to determine the aims that such progress intends to pursue: they are already predefined. In this scenario, answers disappear because some fundamental questions lose their meaning. We no longer ask why this development occurs or why it accelerates so rapidly, for we neither care nor believe we can know. We merely acknowledge that it happens. This is the philosophy underlying the attitude of the major actors driving technological progress because they hold the technologies to do so. As Anderson wrote in 2008, "Google's founding philosophy is that we don't know why this page is better than that one: if the statistics of incoming links say it is, that's good enough. No semantic or causal analysis is required". The causes and semantic relations of progress no longer matter, what matters is only that it occurs. The point is that we are led not to investigate the justifying reasons of this progress, and thus not to seek arguments to confront it critically. As Shoshana Zuboff observes, we have lost control over crucial questions that others have taken over, questions that "define knowledge, authority, and power in our time: Who knows? Who decides? Who decides who decides?"

This so-called digital and algorithmic revolution, far from being a true revolution, is in fact nothing other than the fulfillment of a long process that gained decisive acceleration in the twentieth century. It is a process traced by those who possessed and still possess technology and economic power and who have been able to impose a certain direction in a non-democratic way. The process has experienced pauses in the past, partly due to the practical and ethical-legal difficulties of implementing a fully digitized vision of society. One may think, for instance, of the digitalization of public administration, which still faces both technical problems in transferring analog data into digital form and ethical-legal issues related to achieving so-called transparency.

Despite moments of apparent standstill, the process continued and, from a certain point onward, accelerated rapidly. With the COVID-19 pandemic, this acceleration became extreme. Humanity suddenly found itself projected into a dimension that did not yet globally belong to it. This digital dimension, although it preserved many aspects of life, such as the continuation of work, also revealed its limits, the strain of forcing human relations toward total virtualization under digital surveillance. *Partecipatio* (engagement) in many activities was certainly ensured, but often without the necessary awareness of the nature and pervasiveness of the technologies being used.

In practice, the enormous gap became evident between what technology made possible in a short time during the emergency and the critical preparedness of people to manage it over long periods. This gap does not arise from the intrinsically revolutionary nature of the process, but from the fact that in the planning of technological progress initiated in the last century there was no parallel, equally structured effort to raise levels of knowledge, awareness, and educational preparation among all its recipients.

Since the world became interconnected, first through the Internet, then through mobile phones, and finally through social media, the attitude of those possessing the technology and the economic power to promote it has been to make an ever-growing number of tools available to an ever-wider public, without at the same time creating structural and systematic efforts to build awareness of what those tools really are. One could argue that this responsibility belongs to the public sphere, which should define principles and values to govern technological progress. If this is true, it is equally true that large multinational corporations cannot be regarded as ordinary private actors. They have long exercised the power to delay public intervention, especially through legal regulation, of many of their activities. As Zuboff notes, Silicon Valley's motto has been "innovation without permission". Institutional regulation has been strongly resisted, in order to affirm self-regulation through codes of conduct as the preferred path. However, this conceals the major flaw of self-regulation without a defined legal framework, that it serves only the interests of those who regulate themselves. Such an approach is profoundly undemocratic and unequal. In a complex world like that of technological development, multiple regulatory sources can certainly help, but they must fit within a framework defined by those responsible for ensuring the democratic and egalitarian character of these transformations.

This is why genuine cooperation between private and public actors would be valuable for improving users' capacities for interaction and use, overcoming the dualistic vision of private and public. This is not to advocate the privatization of technological governance. Rather, it means that the principles and rules elaborated at the institutional level must find an ally in their practical implementation by the private actors to whom they apply. In this way, it will be possible to foster a shared vision of the future society, one in which the thresholds incompatible with respect for the human person are clearly defined. Stefano Rodotà already pointed in this direction, but years later cooperation between public and private actors on these issues remains unsatisfactory.

Within the folds of these complex relations between public and private, the dialectic of opposites has taken root, blurring distinctions and intensifying tensions. The anthropomorphized narrative that accompanies human creation of artifacts makes it difficult to maintain the right distance between ourselves and our inventions.

At the institutional and political levels, there have been serious shortcomings in timely initiatives aimed at "providing educational services not only directed at learning how to use certain technological tools but also at the methods that enable a critical approach to them". In the media, however, the narrative of the advent of AI has followed the path traced by cinema and literature, reproducing an anthropomorphizing conceptual taxonomy that reinforces this dialectical

perspective. From this narrative, and from the ubiquity of many notions shared by ordinary, technical, and legal language, stem many difficulties in understanding what is truly happening. If, as we have tried to show, today's technological phenomenon is a profound and largely planned innovation rather than a revolution in the proper sense, then the central terms of the dialectic of opposites (responsibility, person, autonomy, and related concepts) must be redefined to reveal the preeminent role and centrality of the human being in shaping this innovation. This control over meanings will be useful for abandoning substantialisms compromised by the reified assumptions underlying the dialectic of opposites. Since we must now focus on the definitional question, the next section will be devoted to this theme, recalling some key elements of the theory of definition and of the functions of language.

3. *Diffinitio quid nominis* v. *diffinitio quid rei*

In order to analyze a phenomenon through the investigation of the concepts that constitute its main elements of narration, it is necessary to contextualize it within the boundaries of the functions of language, which serve as a prerequisite for the question of definition.

This contextualization makes it possible to highlight the linguistic traps that fuel the anthropomorphizing ideology underlying the dialectic of opposites, since such traps are constructed both around the ambiguous and vague meanings of the terms employed and around the blending of linguistic functions. Through this functional intermixing, one can blur the levels of control over human behavior that language helps to exert.

As was already noted by the pragmatist philosophical orientation, "in all its uses, language performs a function of guiding human behavior". However, one must distinguish between the function of direct guidance and that of indirect guidance. The difference between the two depends on the way language is used in a prescriptive or descriptive function. The first function leads an individual to perform directly the prescribed action, while the assertive-descriptive or cognitive function serves as an indirect stimulus to the performance of actions. The prescriptive function is carried out by propositions that prescribe behaviors by appealing to norms, principles, and values (moral, legal, or social) and thus engages the sphere of the normative. Prescriptions directly promote desired behaviors through orders, obligations, duties, sanctions, and/or incentives. This function reveals itself through judgments of appropriateness or inappropriateness, justice or injustice, regarding the effects of the prescribed behaviors.

The descriptive or cognitive function, by contrast, proceeds through assertions about facts. It belongs to the domain of truth and falsity, that is, to a kind of truth determinable through logical analysis of statements and/or empirical verification. It is within this domain that scientific assertions are situated.

Through the assertions employed by language in its cognitive function, behaviors can be indirectly stimulated. Indeed, by opening, through scientific discoveries and their technical applications, new scenarios for action, the horizons of individual choice are broadened. In other words, individuals find themselves facing unprecedented possibilities for intervention, thanks to the availability of tools previously unknown. One need only think, for example, of assisted reproduction techniques. This important achievement has profoundly transformed the reproductive sphere by making reproduction possible without sexual intercourse. Another example, closer to the theme of this volume, is the advent of the Internet, which has enabled global communication for all, overcoming the problem of physical distance.

The possibilities of action offered by science and technology are not limited to having utilitarian consequences in terms of opportunity. Underlying these new scenarios are value choices that compel reflection on the ethical sustainability of new opportunities and on the need or not to establish specific conditions for their realization.

When we speak of value choices, we must be clear about what we mean. One must not fall into the error of believing that the descriptive and normative domains represent a kind of continuous line along which prescriptions can be derived from descriptions. This type of conclusion has been the subject of deep criticism in the debate surrounding the so-called is-ought question, that is, the thesis of the "Great Division". The debate concerning the ontological background of the distinction between language functions has been at the center of intense controversy, especially during the twentieth century. As Uberto Scarpelli reminds us, for some, the distinction corresponds "to a division of reality, and of the experience humans have of it, into two dimensions: the dimension of being and the dimension of ought", while for others we are instead dealing with two different uses of language "within the single and continuous context of human experience". This is not a merely theoretical or academic debate. On the contrary, it has significant practical implications for the transparency of underlying ethical positions.

Indeed, adhering to the thesis of the Great Division between the descriptive and the normative means taking a stand for discursive transparency. It begins with the recognition that it is not possible to move from the descriptive function of language to the prescriptive one by purely logical-linguistic means, without committing a logical leap. Those who support this view consider it possible to construct transparent discursive universes, capable of argumentative rigor and open to intersubjective verification, without confusing this method with the experimental method proper to strictly scientific contexts.

The debate on the is-ought question, in its many variations, is far from resolved, since no definitive consensus has been reached concerning the scope of the Great Division between the two functions of language. What does appear to be established, however, is that maintaining this distinction better satisfies the need for transparent discourse regarding premises and the conclusions derived from them, by means of the verification of intermediate steps, that is, through logical and/or empirical control.

It follows, therefore, that even though the confrontation between divisionists and anti-divisionists remains open, the practical consequences of belonging to one camp or the other are clear. Holding firmly to this distinction allows us to address certain critical novelties of current technological development, which can apparently give new life to this long-standing debate.

As will be analyzed more thoroughly in the next chapter, some of today's technological developments challenge the aforementioned distinction between the direct or indirect influence of scientific and technological progress on human behavior, and consequently make it difficult to defend the separation between the two functions of language. Some technologies, in fact, can directly affect conduct, inhibiting or promoting specific actions aimed at a particular goal. This is the case for those technologies that allow externally directed control of devices and instruments belonging to an individual, for example, when someone who fails to pay car insurance can be prevented from starting the vehicle. Faced with this novelty, one may ask whether the is-ought question can truly be regarded as resolved in favor of those who "deny the Great Division because it divides too much", or whether there are still valid reasons to continue defending this distinction. Those who take the latter position emphasize the need for transparency in presenting arguments and positions on issues of great transformative significance, as is the case with today's technological innovation. Distinguishing between facts, operations, and the discourses

surrounding them represents a form of control over the ends and values underlying the direction of society's digital transformation. Such transparency is all the more necessary in the many situations, both in ordinary and institutional language about AI, in which it becomes difficult to distinguish between the cognitive and the prescriptive function. This difficulty arises because the two functions of communication do not always operate independently and are not always easily distinguishable. Often the terms used, or the way words are combined in statements, conceal a persuasive dimension that appeals to emotional aspects, making it difficult for the recipient to discern the informational content from the prescriptive elements of the message.

In the narrative surrounding AI, the blending of different functions and uses of language (often prompted by persuasive intent) constitutes one of the greatest sources of difficulty, especially for the final recipients of digital transformation, namely, citizen-users. When AI is discussed, not only in the media but also within institutional debates, it is not always easy to distinguish between the cognitive and the prescriptive function. There are, indeed, several discursive situations, some of which may serve as examples.

There are cases in which it is relatively easy to confine the discourse within one of the two language functions. This occurs, for instance, when the purpose is purely informative, when one intends to describe, for example, the potential applications of AI-driven technologies or recall how certain results were achieved. Equally clear is the case when one writes in explicitly prescriptive terms, indicating the need to pursue a certain direction for its development. For instance, when Ursula von der Leyen, President of the European Commission, stated that "[…] we must now make this decade Europe's Digital Decade so that all citizens and businesses can have access to the best the digital world can offer. The Digital Compass presented today clearly shows us the course to follow in order to achieve this goal".

Much more often, however, the narrative becomes charged with triumphalistic and evocative tones, with "euphoric" values. In this way, the persuasive dimension merges with the descriptive and/or prescriptive one. An example is the statement by Margrethe Vestager, Executive Vice President for a Europe Fit for the Digital Age, who declared: "Together with the European Parliament, the Member States, and other stakeholders, we will work to ensure that Europe becomes the prosperous, determined, and open partner we want it to be on the global stage, and that each of us can fully benefit from the well-being generated by an inclusive digital society".

Once the distinction between the functions of language has been drawn, attention must be turned to the issue of definition. This issue concerns not only the separation between the prescriptive and the cognitive function of language but also the distinction between different conceptions of language, particularly between the conventionalist and the essentialist views. According to the former, language is a cultural institution, a means of communication among members of a community, expressed through words (artificial symbols) combined into meaningful statements. On such a conception of language, linguistic control is possible through adjustments of meaning, which occur over time with the introduction of new words or with the revision of existing meanings in light of social and cultural change. From this perspective, it becomes relevant to determine the conditions of use of a term in relation to the aims one seeks to achieve. In ordinary language, this is determined by the established patterns of use among speakers of a given linguistic community.

Unlike this approach, the essentialist conception of language begins from a specular relationship between reality and words. In other words, between words and what they signify there exists a natural relation, inherent to the order of things. The language user cannot performatively act upon the meaning of words but can only discover and acknowledge it. In this view, there can

exist only one true meaning for each term, reflecting its essence and therefore predetermined by reality.

The two conceptions of language, briefly recalled here, influence the use of definition in different ways. As Herbert Hart explained, "definition […] essentially consists in drawing lines and distinguishing between one kind of thing and another, which language designates by separate terms". If one follows the conventionalist perspective, the definition, called nominal definition, concerns words, and three types are distinguished: lexical, stipulative, and explicative or re-definitional.

A lexical definition describes the use of a term within a given group of speakers. A stipulative definition creates new terms or radically modifies the meaning of existing ones by stipulating a specific sense. An explicative or redefinitional definition serves to delimit or clarify the meaning of a term functionally to the objectives of the discursive universe in which it is employed, while maintaining a certain link with ordinary usage. In the conventionalist approach to language, definition thus becomes an instrument for bringing order through control over meanings. It becomes a means of overcoming certain problems already noted by John Austin, namely: (1) if words are tools, we must strive to use clean tools so as not to fall into linguistic traps; (2) since words are not facts or things, we must be able to abstract them from the world in order to recognize potential inadequacy or arbitrariness and thereby address reality without blinders. The definitional instrument allows precisely this operation: by delimiting meaning, we may hope that disputes apparently irresolvable because of linguistic misunderstanding will find their way toward the sharing of practical solutions.

The role of definition changes completely if it is related to the essentialist conception of language. Its object, nature, structure, and function are transformed. In this perspective, the definition, called real definition, concerns things. It may be direct or formulated by genus and specific difference, it is resolved into a statement that is true or false, and has a cognitive function, because it enables us to describe the essence of the things of which words are the mirror. Because meanings are predetermined, it is not possible for speakers to modify or redefine them. The real definition and its associated essentialist conception of language lie at the root of some of the most pernicious hypostatizations that populate the dialectic of opposites today. Although both the essentialist conception and the theory of real definition have been subjected to incisive critique, they remain subliminally widespread, even among professional categories, and are thus even more deeply rooted in the minds of ordinary people.

Just as the functions of language do not always operate autonomously, and it is difficult, without adequate linguistic analysis, to separate the descriptive plane from the prescriptive one because persuasive intent often creeps between the two, so too can the definition of terms take on persuasive overtones. Through persuasive definition, one seeks to validate as certainties what are in fact beliefs grounded on subjective bases and functional to undeclared interests. Persuasive arguments can rely on the persuasive definition of certain terms whose meanings are not univocal and which, over time, have acquired an emotional connotation, either extremely positive or extremely negative. Among the terms that have undergone this fate, one may recall, for example, eugenics. Its lexical meaning refers to the branch of genetics devoted to the progressive improvement of the human species through the mating of individuals bearing favorable genetic traits. However, following the discriminations and abuses perpetrated under the guise of misinterpreted eugenic policies, the term acquired a strongly negative persuasive meaning from which it has never been freed.

Persuasive connotations, together with the persistent essentialist conception of language, which equates the meaning of a word with an alleged truth, inducing the belief that concepts and categories (including legal ones) are "fixed entities, given once and for all and indisputable, applicable in all fields", constitute an integral part of the narrative of AI and serve the interests of those seeking to impose a single path, that of so-called *algorithmic governmentality*, without offering alternatives.

This is why it is of great practical importance to determine as clearly as possible the meaning of the terms used in the various discursive contexts surrounding AI.

Clarity therefore requires, first of all, the definitive abandonment of the essentialist conception of language, which unfortunately still underlies not only common debates on AI but also, and even more gravely, legal discourse itself.

If the process of determining meanings is viewed through the lens of the essentialist conception, one will tend to overlook the volitional and decisional component that lies behind the use of certain terms, since emphasis will be placed instead on a so-called intrinsic or proper meaning of the defined term. This approach has enormous practical consequences. It allows those who possess not only technological but also economic and communicative power to perpetuate an anthropomorphized representation of AI, one that fosters an illusory trust in technologies rather than highlighting the issue of trust in those who design, produce, and market them. Moreover, it enables those in power to carry out subtle and pernicious operations, such as: (1) mystifying the course of technological progress through the rhetoric of inevitability; (2) obscuring the ethical choices underlying technical and economic decisions; and (3) evading responsibility as communicators of accurate and transparent information about the scope of innovations they make available to individuals and to society at large.

In sum, much of what has been harmful to the integrity of democratic societies and to the values of equality and non-discrimination in the context of technological development has occurred through language, particularly through choices that have favored semantic ambiguity and vagueness over rigorous linguistic control at the communicative level.

It is undeniable that governing language is not easy in a society characterized by the exponential speed of communication. Speed, together with the proliferation of information channels alongside institutional ones, not only prevents those responsible for informing from always exercising the necessary care for linguistic accuracy but also discourages audiences themselves from demanding it, as they increasingly prefer flash news to articulated explanations. The problem is not only grammatical or syntactic accuracy. The real issue is that of informational illusion, since the recipient is regarded more as a target to be captured than as a subject to be informed. In this context, the persuasive dimension plays a central role.

Its silent presence undermines the communication of information on complex topics with strong implications for the lives of recipients. Communicating technical matters to the public does not mean using difficult language, but avoiding fake simplicity, that is, words that appear simple but, through overuse, have lost their informative power. To enable recipients to form their own judgments on issues that affect their lives but are highly technical, one must not aim for "easy speech" in the sense of banal simplification, but rather pursue clarity, the kind of clarity that enlightens, that is neither deceptive nor complacent, neither hasty nor reductive.

This is certainly an ambitious goal, but one worth pursuing, insofar as we wish to be individuals who play their part in the historical moment they inhabit and, in doing so, shape the

conditions in which future generations will live. To achieve discursive transparency and clarity, it is necessary to reflect on the concepts used in the narrative of AI, starting from the recognition that those who label it as a "revolution" are engaging in a persuasive operation that relies on the vagueness and ambiguity of many other terms that have accompanied the development of intelligent technologies. We may thus speak of a conceptual taxonomy of the anthropomorphic vision of AI.

4. Conceptual taxonomy of the anthropomorphization of AI

The anthropomorphism of machines has for decades been critically addressed in literature. Frank Pasquale highlights the danger of an anthropomorphism he calls deceptive, realized through children's exposure to robot teachers as a form of "subtle indoctrination to teach that human and machine are ultimately equal and interchangeable". Anthropomorphization has many forms and can persist through various channels. Its central point, which allows it to be pervasive in discourse, lies in the use of specific concepts in the narration of the history of AI. The conceptual taxonomy of this history nourishes the dialectic of opposites and makes it difficult to bring order to this narration. To reach the root of the problem, it is necessary to cut through the dense network of anthropomorphizing terms. The terms chosen for semantic analysis have these characteristics: 1) they can be used ubiquitously in different discursive contexts, from ordinary to specialized language; 2) they are normative and lack a semantic referent in reality; 3) they have acquired strong persuasive value through use; 4) they have historically referred to man to indicate distinctive qualities, especially in relation to free will. The terms that meet these conditions, (autonomy, person, responsibility, unpredictability, and intelligence) form the basis of the process of anthropomorphizing machines and support the human-machine dichotomy. All have an ancient semantic history and are mostly linked to the traditional debate on freedom of action, namely on free will. Only intelligence did not explicitly belong to that long-standing dispute between the defenders of free will and the supporters of a deterministic and reductionist view of human action. However, this term forcefully enters the current dialectic of opposites and has perhaps favoured its wide diffusion. The whole history of AI, since the expression artificial intelligence was coined in 1956 at the Dartmouth Summer Project on Artificial Intelligence, the founding moment of cybernetics, revolves around defining AI *per relationem* to human intelligence. One of the main figures of this development, Alan Turing, already showed this tendency in his 1950 article Computing Machinery and Intelligence, where he examined the question of creating thinking machines endowed with a human characteristic and did not exclude such a possibility.

The notion of intelligence is therefore of crucial importance in the narration of AI. It belongs to the grid of primary concepts through which the dialectic of opposites and the anthropomorphism of technology have taken shape. These concepts firstly designate qualities of actions and/or of the agent. One says, for instance, "he acted autonomously, responsibly, unpredictably, intelligently", or "she is a responsible, unpredictable, autonomous, intelligent person". These are terms with an ancient semantic history, and their use requires clarification of the conditions of use, that is, the classes of things or facts named by the word (extension) and the properties that must be shared for inclusion in that class (intension). Such definitional operations, proposed for the selected terms, are fundamental for understanding the dialectic of opposites.

4.1. Intelligence, autonomy, and unpredictability of the human person

Let us first consider these terms when they are used with reference to the human being, beginning with intelligence. In common understanding, the word intelligence refers to the possession of several properties, ranging from those that are strictly rational to those that draw upon emotional aspects. Intelligence is often qualified by adjectives, as in the expressions "practical intelligence", "speculative intelligence", or "rare intelligence", "acute, weak, slow intelligence", and so on. The range of mental and psychic faculties attributed to the human being, which we summarize under this term, is therefore very wide.

If we wish to delimit the minimum conditions of its use, we may recall the psychic and mental capacities that allow one to reason, to understand reality, and to situate and orient oneself within it. However, human intelligence is not limited to such mental faculties, but also includes corporeality and emotionality. The richness of the word's semantic nuances, together with its positively persuasive connotation, may easily give rise to misunderstandings if those who employ it do not redefine the conditions of its use within a given context.

When the expression *Artificial Intelligence* was coined, little reflection was devoted to the implications of that linguistic choice nor was it imagined that it would foster a dialectic of opposites and a profound dichotomy between the human and the mechanical. The name chosen in 1956 for these new machines triggered an anthropomorphizing ideology which can today be exploited by those who wish to make machines appear superior to human beings (what I shall call excess anthropomorphism).

If the notion of intelligence is the primary concept of the narrative surrounding AI, then it deserves institutional attention and an adequate redefinition. This indeed occurred with the work entitled A Definition of AI: Main Capabilities and Disciplines (2019), elaborated by the High-Level Expert Group on Artificial Intelligence. The document sets out to define AI in order "to avoid misunderstandings, to achieve a shared common knowledge of AI that can be fruitfully used also by non-AI experts, and to provide useful details that can be used in the discussion on both the AI ethics guidelines and the AI policies recommendations".

The Group begins precisely from considerations regarding the vagueness of the term intelligence and seeks to delimit its meaning. It recognizes that the content of intelligence has been the object of study across various disciplines, and concentrates on the fact that, in the field of AI research, the term refers to rationality. The Group adopts the definition elaborated by the European Commission, according to which: "AI refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals". The Group further specifies that an AI system is: "Any AI-based component, software and/or hardware. Indeed, usually AI systems are embedded as components of larger systems, rather than stand-alone systems". It then explains how such systems reach this form of rationality: "By perceiving the environment in which the system is immersed through some sensors, thus collecting and interpreting data, reasoning on what is perceived or processing the information derived from this data, deciding what the best action is, and then acting accordingly through some actuators, thus possibly modifying the environment. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions". Finally, the Group proposes an updated definition: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information derived from this data, and deciding the best action(s) to take to achieve the given goal. AI systems

can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions".

This institutional redefinition represents an important step toward clarity and transparency concerning the potentialities of AI. From it, we see that AI is a human-made system able to carry out a complex task by using sensors to investigate the environment and actuators to act upon it. Absent from this definition is any reference to emotional intelligence. The focus is instead on the connection between intelligence and rationality. In this way, the sense of intelligence as rationality is accentuated, that is, as the ability to act "according to the efficacy of the means chosen in view of an end". Such a redefinition thus highlights one aspect of intelligence that can be shared by both human beings and machines, while indirectly allowing a revaluation of human intelligence in the light of further criteria that distinguish it. For although in common language rationality is often used as a synonym for reason, and rational as a synonym for reasonable, in philosophical terms the two concepts have quite different nuances. The notion of reason, though it too admits of several meanings, has a core sense defining it as "the autonomous guide of the human being in all fields in which inquiry or research is possible" and, as such, "the force that liberates from prejudice, myth, and false opinions, and that allows the establishment of a universal or common criterion for human conduct in all fields". By redefining intelligence as rationality rather than reason, the scope of the term is considerably narrowed. One can therefore predicate of AI an intelligence understood as rationality, but not as reason. AI may be said to possess a specific rationality, determined by actions based on data collected and interpreted through predetermined interpretive frameworks, with a degree of operative latitude that is, however, decided in advance by those who design the system. At least at the current state of technological development, it is not an intelligence that can also be reasonable, that is, based upon reason in the sense just recalled. Clarifying this point provides a means of countering the representations of AI as intelligent in excess (excess anthropomorphism), and allows for a renewed appreciation of human intelligence, one that restores it to the centre of attention not as something to be plundered in order to grant machines a similar quality, but as something to be cultivated, not only to improve intellectual performance but to ensure the pluralism of its features and, with it, equality among human beings, marking the difference in excess in favour of the human and not of the mechanical.

The human being is, therefore, endowed with reason, to be understood as the force at the foundation of judgment and discernment. With reason, fully developed in the Kantian sense, the human being can attain autonomy and freedom.

We thus arrive at one of the most controversial concepts employed in this dialectic of opposites, namely that of autonomy. The term appears frequently, especially when reference is made to self-driving cars or to autonomous robots capable of self-learning and therefore said to be unpredictable in their actions. The concept of autonomy has a long history in philosophical thought and has, for several decades, stood at the centre of debate on ethically significant questions such as those concerning end-of-life decisions, procreative domains, genetics, and the neurosciences. In its semantic history, the meaning of autonomy has been applied exclusively to human beings and is closely linked both to freedom from constraints and external impediments, and to freedom to, that is, the capacity and power to make choices. This distinction was treated by Isaiah Berlin, who, in distinguishing the two senses, clarified that freedom from (negative liberty) designates "the area within which the subject is or should be left to do or be what he is capable of doing or being, without interference by other persons". In its second sense, freedom to (positive liberty), autonomy manifests itself as the condition of being master of oneself, that is, the power to make autonomous choices.

Both meanings of autonomy include, in turn, a reference to responsibility and to unpredictability. It follows that if a subject is autonomous in making choices, they must also bear the consequences of those choices for others and will therefore be held responsible. A human autonomous decision is a decision taken all things considered, that is, one that respects certain conditions: the absence of physical or psychological coercion; the presence of a capacity to understand a range of available and relevant information such that the decision is not manipulated; and the ability to act on the basis not only of considerations of utility or expediency but also through the integration, in the deliberative process, of a balance between values expressed by moral and/or legal norms that guide one's own and others' behaviour. When we speak of value choices, we refer precisely to this. The values guiding decisions may change over time, and thus the choices made may later be reconsidered in light of new evaluations. Autonomy therefore finds expression in the very possibility of changing one's position, while assuming responsibility for it. Such responsibility may be moral, legal, political, or social. The concept of responsibility always entails reference to norms of conduct in its definition. It is a normative concept, in which there are always elements that can be called artificial, in the sense that they are socially determined. As Uberto Scarpelli clearly explained, the use of this concept requires that three elements always be present: a duty, a consequence (sanction), and a behaviour.

Whoever is responsible for their actions is a moral agent or person. Their responsibility with respect to the duty to be observed may be modulated depending on whether pathological impediments exist, or situations arise that require assessment of their capacity to act autonomously. Modern legal systems presume full capacity in every agent or person, providing exceptions only in particular and demonstrable cases. It is therefore at the level of the consequence that the law operates when it is proven that a person lacked full capacity, not by modifying the duty prescribed by the norm of conduct, which remains unchanged. Through this schema of responsibility, a functional relation is established between the subject and the actions they perform. In practice, this schema of responsibility is a social construct, that is, "a socially established link related also to cultural criteria". Underlying this schema lies the pursuit of a social order in which individual conduct is oriented in such a way as to limit or avoid unpredictable behaviour.

The unpredictability of human actions can be interpreted in two opposite senses. In the first sense, an unpredictable subject may be one who is out of control, who engages in behaviour that violates shared norms of conduct. In the second sense, a subject may be unpredictable in that they are capable of surprising others through uncommon, though not uncontrolled, ideas and actions. In this latter case, the regulative schema of social conduct is respected, but the behaviour enacted is not that which one would ordinarily have expected. When a subject is out of control, the reasons for such conduct must be investigated, and it must be verified whether the person can in any way be held responsible for violating the socially established regulatory schema. When, instead, a subject acts in an uncommon way, there is no need to investigate the reasons for their action or their mental state, for their behaviour is entirely consistent with the shared norms of conduct, and they are therefore fully responsible.

Moral and legal responsibility for one's actions, and thus subjective autonomy, is the reflection of a capacity to act otherwise. In twentieth-century philosophical literature, much attention has been devoted to responsibility and the freedom to act, often through analysis of the proposition "he could have acted otherwise". In examining this proposition, Alf Ross emphasized that defining whether one could have acted otherwise depends on the coexistence of three conditions: constitutive or constitutional conditions (that there be no constraints such as mental illness), occasional conditions (that the circumstances allow the action), and motivational conditions (that represent the will or motive for the action). According to Ross's analysis, given the constitution and the occasion, when we say that a subject "could not have acted otherwise" in

situations of coercion, we are in fact saying that "we could not expect him to have acted otherwise, although he could have". The emphasis thus falls on the will and on the capacity to decide autonomously.

From both the moral and legal points of view, the exercise of autonomy and the attribution of responsibility presuppose a moral agent or person. They therefore call into question the notion of person, which, like the other fundamental concepts discussed, has a long semantic history during which multiple meanings have become layered. It is a normative concept that refers to the normative treatment corresponding to the ascription of rights and duties. The person is not the individual in a biological or psychological sense, but rather the normative treatment, or, if one prefers, the set of norms that, at a given historical moment, are established in favour of or against certain lines of conduct. This explains why not all those who, in a descriptive sense, belong to the human species have always been treated as persons in the moral or legal sense. One need only recall slavery, or the fact that the unborn is not a person in the legal sense, though it is a subject of rights. The separation between law and nature, and between law and morality, which resulted from the legal positivist reflection of the past century, contributed decisively to freeing the concept of personfrom factual referents. To this was added another essential step, the result of modern constitutionalism (the so-called constitutional personalism), which connected the notion of person to equality, not only formal but also substantive, thus overcoming "the problem of compatibility between the abstract subject and the recognition of differences". Equality in rights allows equal value to be attributed to differences and constitutes the value or dignity of the person.The current notion of person and its use in contemporary legal contexts therefore presuppose specific value choices. Through this term, one takes a stance in favour of an inclusive and anti-discriminatory vision of society. It is not necessarily a notion that must be pre-conceptually closed to the inclusion of other entities, such as nature or AI. What the current ethical and legal meaning of the term person indicates is that its extension also entails the extension of the egalitarian dimension and this must therefore be carefully considered whenever one proposes to include other entities.

From what precedes, it emerges that the notions analysed in this paragraph have been historically referred to the human being, not because there exists any "true" definition, but because, regardless of the nuances their meanings have taken on through time and use, these words have come to recount the conceptual evolution of human behaviour in society.

Starting from these premises, it becomes necessary to clarify how these same notions are employed in the narrative surrounding AI, and in what way they sustain the dialectic of opposites and the anthropomorphic vision of technology.

4.2. Intelligence, autonomy, and unpredictability of the artificial person

The European institutions have intervened to define what is meant by intelligent machines, with the aim of clarifying and making understandable, even to non-experts, the realistic prospects for applying what we call intelligent machines. In view of this important institutional effort, which will hopefully have positive effects on the dialectic of opposites that pervades the media narrative of AI, what must now be done is to ensure that this linguistic control finds systematic implementation, so as to correct the course toward a discussion and a narrative of AI free of gross misunderstandings.

In this direction goes Kirchschlaeger's proposal to replace the term Artificial Intelligence with data-based systems. This is convincing insofar as it "lifts the veil of the inappropriate

attribution of the myth of 'intelligence,' which covers substantial problems and challenges of data-based systems, and allows for greater accuracy, adequacy, and precision in their critical reflection". The author maintains that this expression better explains what AI actually consists of, since it emphasizes processes of data generation, collection, and evaluation from which derive data-based perception, prediction, and decision. In short, data-based systems directly and explicitly recall the central node of AI, which is the capacity to possess and use an immense quantity of data. If for this notion there already exist institutional and theoretical proposals for redefinition, the same cannot be said for the other primary concepts used in its narrative. To prepare the way for their redefinition, we propose some reflections on autonomy, unpredictability, person, and responsibility when they are used with direct reference to intelligent technologies.

As regards autonomy, we have seen what its semantic range includes when the discussion concerns the human being. Here, the question arises of what is meant when we speak of autonomous robots, autonomous vehicles, or autonomous AI.

The use of autonomy in relation to intelligent technologies recalls both the power to act without external impediments in the physical space in which AI operates and the power to choose that action which, according to a calculation of advantage or risk-benefit, best achieves the goal pre-programmed by humans. The discriminating point between human and machine action lies precisely in how we understand the power of choice. The choice of AI is a weak choice: although AI performs the risk-benefit calculation, the aim and the purpose for which the calculation is made are heteronomously imposed by humans. By contrast, human choice is strong, based on deliberation guided by internalized values that inspire judgment on one's own and others' actions. When the goal is heteronomously set, as in AI, so too are the values underlying its pursuit. In the case of autonomous driverless cars, the modalities of decision in morally charged situations, such as the well-known trolley dilemma, are predefined by the programmer, and AI can only select among the options made available to it, each already embedding moral choices. Unlike human decisions, autonomous in the strong sense, the deliberation based on non-pre-programmed values remains excluded from machine decisions. If this fundamental difference is neglected, there is a risk of conveying the idea that human and machine autonomy are equivalent, an operation detrimental to the appreciation of the moral complexity that characterizes human action. The latter involves intricate deliberative processes, the result of a long and elaborate development of individual moral consciousness. These processes incorporate various criteria that contribute to the final decision in ways not yet fully known. Emotions and sentiments, as well as beliefs and conative states, all play a role, as already argued by philosophers such as Mario Calderoni. According to Calderoni and as contemporary thinkers like Shaun Nichols and Hanno Sauer confirm moral judgment arises from the joint contribution of emotion and reason. Excluding pathological conditions, human beings can act otherwise, both because there is a creative unpredictability in their actions and because they are free to fall. This last component is precluded to AI, unless one intends to program it to do harm. When unpredictability is attributed to AI, it refers to a technical unpredictability, not to self-determined deliberative choices. The unpredictability of machines originates heteronomously, from programming choices, from the values and action criteria introduced, and from those excluded.

Clarifying these semantic differences is a way to dismantle the dialectic of opposites not only in communication and media but also within specialized contexts where it has unfortunately taken root. The ubiquity of the primary concepts analysed so far has facilitated the spread of the anthropomorphizing ideology of machines. Certain contexts, by their nature, even risk perpetuating, indeed ennobling, this toxic anthropomorphism. This is the case of law, where, since ancient times, the contest has been played between reductionist visions of the human and those aimed at affirming freedom.

To conclude our analysis of the conceptual taxonomy implied in the narration of AI and in the underlying anthropomorphizing ideology, it is necessary to recall the last two primary concepts, namely, responsibility and person. The question is whether AI can be considered responsible for the actions it performs, for example, through robotic devices, and therefore whether it can be qualified as a person, that is, a moral agent.

It is precisely the association with machines through the notions of autonomy and intelligence that creates high expectations regarding the capacity of intelligent systems to behave like persons, that is, moral agents. As has been observed, "because of their increased intelligence, autonomy, and interaction capabilities, AI systems are increasingly perceived and expected to behave as moral agents". These expectations depend in large part on an intrinsic predisposition of human beings to consider as human-like machines with certain characteristics, since "[e]ven if you know a robot has very little autonomy, when something moves in your space and it seems it has a sense of purpose, we associate that with something having an inner awareness or goals". The concept of person represents a formidable pass for entering the world of norms, rules, rights, and duties that govern the context of human relations. The definition of who is a person and who is not has practical implications for the distribution of benefits: it is a battle for power. Being persons in the sense of moral agents implies having to / being able to also assume moral responsibility for the actions one performs. It means stepping onto the highest step of the podium, the one on which, until now, only human beings have been placed. It is no small matter to think of and choose to define AI as a person. With such a choice, which has strong moral and legal implications, the traceable boundaries between us and the fruit of our creativity are erased.

As we will see in the chapter dedicated to legal issues, what is at stake is the very stability of fundamental rights, historically not by chance called droits de l'homme.

In short, in light of the considerations developed in this paragraph, the force of the semantic question should have emerged, in determining the direction that the relationship between power and language may or may not take, and how control over language becomes an essential instrument of defense or, alternatively, of contempt for human dignity.

5. The semantic issue as an ethical problem

From the preceding semantic analysis, we can affirm that the narrative of AI, in all its variations and nuances, has never been neutral, that is, detached from its sociocultural context, from the pressures of economic power, and from the values underlying the political and legal framework in which it develops. Even its descriptive part, produced by technicians and experts, is itself a product of the historical and cultural context in which scientific progress takes place. On this point, Sadin has clearly noted the singular nature of the historical narrative that characterizes the evolution of informatics, observing that "a kind of grand narrative has taken shape, always repeated in the same way, which creates a static framework" incompatible with the irreducibility of human experience, which history usually recounts. The history accompanying the evolution of AI therefore appears decontextualized and has nourished the belief in the neutrality of technology. But the very concepts used in its narrative most clearly show the non-neutrality of what is taking place.

If this narrative is not neutral, the problem lies not in language itself but in its use. It is a strategic use by the main actors of the narrative to present a story to the public. It is, therefore, a matter of choice.

Just as there exists a narrative shaped by the dialectic of opposites and by anthropomorphizing ideology serving a few holders of power, another and more plausible narrative is possible. Frank Pasquale's analysis shows that a different path for the development of robotics and AI can be outlined through courageous choices based on three premises: to support the complementary, not substitutive, role of AI with respect to human labor; to avoid radical transformations in all sectors and maintain the status quo in some; and to recognize the institutions capable of pursuing these aims. These premises are entirely shareable and should be realized not only through the author's proposals but above all through control of the language used. Without such attention, it becomes difficult to resist the commercial and entrepreneurial logic that seeks total automation of human activities. Language produces habituation, and habituation weakens the impulse to change direction.

One of the main factors that has allowed these strategies to spread is the persistent and apparently unassailable vision of a future of well-being and prosperity achievable mainly through technology. Language has played a crucial role in presenting this project. Consider the opening of the White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (19 February 2020): "Artificial intelligence is developing rapidly. It will change our lives by improving healthcare (for example by making diagnoses more precise and allowing better prevention), increasing agricultural efficiency, contributing to climate change mitigation and adaptation, improving production efficiency through predictive maintenance, increasing the safety of European citizens, and in many other ways we can only begin to imagine". A brief mention of possible risks follows, but the initial emphasis on positive aspects already reveals a clear orientation.

Although the European approach is gradually distancing itself from the previous dialectic of opposites and aims at AI oriented to responsibility and respect for rights and freedoms, it still shows contradictions and ambiguities often linked to language. Fragments of the earlier narrative persist beneath institutional discourse, supporting the idea of a technological revolution as a radical break with the past and promoting the image of a technology able to solve all human problems. Within institutional discussion, objectifying and absolutizing representations of AI remain, influencing also how data is conceived. Data is often treated as a pre-social given, when in fact it is the result of social and cultural construction based on ethical choices.

The so-called data-driven vision of society, in which data is used for decentralized and supposedly more efficient decision-making, is often confused with a model where AI gives objective and definitive answers about reality. This happens because it is contrasted with the so-called human-centric model. Once again, the dialectic of opposites reappears between a technology centered on human beings and one acting autonomously and building its own interpretation of reality from the data it processes. In the confusion generated by this opposition, it becomes difficult to grasp the subtle but essential difference between a data-driven and a human-centric society. This difference concerns the distribution of the benefits of technological transformation. The human-centric model explicitly adopts an egalitarian perspective, according to which all people should benefit from technological development. The data-driven model is less clearly inspired by this ideal and more likely to obscure who actually benefits, namely those who hold technological and economic power.

Unmasking the false dichotomy between a data-driven and a human-centric society allows us to see the ethical relevance of semantic choices. The question of AI is above all a question of power and its control. If this is true, semantic precision becomes a condition for the preservation of the rule of law and democracy. It therefore concerns institutional, political, and legal language, because of its effects on fundamental rights and the stability of democratic systems. The semantic question, or the control of language, is therefore profoundly ethical.

The model of a society that finds its answers in algorithmically interpreted data is based on what is called technological determinism. This concept, well known in the debate on scientific progress, particularly in genetics, holds that technological development shapes society autonomously and inevitably, independent of the social context, and that it is neutral and unavoidable. But this view confuses facts with processes. Technological development is not a single event that happens once and for all. It is a changing process, deeply influenced by economic and power structures, and by the culture and values in which it occurs. If we distinguish the descriptive level from the prescriptive one, it becomes clear that technological progress is not a natural given but the result of human choices aimed at building intelligent machines, despite difficulties and obstacles. This process began with the Dartmouth Summer Project on Artificial Intelligence in 1956. When this distinction is not maintained, as often happens in media and institutional discourse, value-laden decisions are presented as neutral facts, leading to uncritical acceptance of technology and of the power that governs it.

The history of AI shows that, from the beginning, it has been the object of attention and planning by powerful actors, such as the Rockefeller Foundation, who influenced political decisions and investments. The continuity of this process is not due to natural evolution but to human choices. If the technological transformation had truly aimed to be democratic, its guiding principles should have been explicit from the outset. Instead, the prevailing attitude, often institutionally supported, was to conceal them behind a prescriptive narrative disguised as descriptive. Transhumanist philosophy reinforced this approach. Though not a unified movement, transhumanism generally supports the use of technology to enhance the human being without particular ethical questioning. It rests on the idea that everything technologically possible should be used. In this view, human limits are seen as problems to overcome, indirectly supporting the strongest form of technological determinism. However, technological development is not neutral or self-justifying. It is the result of moral, political, and social struggles over which directions should prevail. Emphasizing the relation between power and technology does not mean denying the internal logic of experimentation, which guarantees the validity of innovation. But the logic of scientific control differs from the context of invention, which situates technological development within specific historical, cultural, and political conditions. The notion of context of invention, taken from the philosophy of science, helps explain the paths followed by technological development and the choices made at certain moments. According to Scarpelli, what belongs to this context includes preferences and initiatives that are not objective but shaped by the surrounding environment. As such, they are subject to moral and political evaluation, like all social activities, in relation to their origin, development, and effects. Maintaining the distinction between the context of control and the context of invention makes clear that scientific and technological progress is neither predetermined nor inevitable. Awareness of its non-neutrality opens the way to a critical evaluation of its directions and choices. Its pervasiveness requires the participation not only of those who produce and commercialize technology but also of those who are subject to it. Following this line, a further step is necessary, which is to make participation effective. This means not only formally recognizing autonomy, as many transhumanists do, but creating the conditions for its realization. People must be enabled to make informed choices, with clear tools of information, questions, and answers. Autonomy is not innate but a process that must be cultivated and supported by those who hold informational and communicative power.

We thus face a linguistic trap. There's the affirmation of autonomy only in form, underlying the imbalance of a narrative marked by excessive triumphalism around AI. As Frank Pasquale observes, the balance between humans and machines is changing in daily life, and avoiding the worst consequences of this transformation while realizing its benefits depends on our ability to act on this balance. Achieving it requires a narrative that finds the right measure between enthusiasm and rejection. This balance is possible if we abandon dogmatic attitudes disguised as neutral

descriptions and respect the difference between what can actually be done with current technologies and what is only projected or promised, confusing what is with what might be.

As George Orwell reminds us, the great enemy of clear language is insincerity, which uses euphemisms, vague precision, and false arguments to conceal the distance between real aims and declared intentions. The value of a linguistically rigorous and semantically clear approach, which maintains the distinction between descriptive and prescriptive levels, is also evident when attention shifts from transhumanism to bioconservatism.

Although this current contains internal differences that will not be addressed here, some shared assumptions can be identified. A precursor of these ideas is Hans Jonas, who writes that grounding value in being itself means overcoming the supposed separation between what is and what ought to be. If what is good is so in itself, it carries the demand for its realization, becoming a moral imperative. Starting from the idea that human nature follows the natural order of things, bioconservatives maintain that it must always be preserved. Technology is seen above all for its harmful potential, as an instrument that risks breaking with this natural condition, except when used for therapeutic or care purposes.

For this reason, bioconservatives are favorable to the therapeutic use of technology and contrary to uses beyond therapy. Despite their differences, transhumanist and bioconservative positions share an ideological nature marked by dogmatism, objectivism, and essentialism. The result is a tendency to limit the use of technologies according to their perceived invasiveness, based on the belief in a final good embodied in an ideal of virtuous life.

From what has been said, it should be clear that extreme positions are unsustainable. The question is not to demonize or glorify the machine but to make choices guided by freedom and transparency, supported by ethical awareness, and directed toward the common good.

6. The semantic issues as a constitutional and democratic question

For *partecipatio* [engagement] in the digital transformation of society to be truly effective and productive for all, the information circulating must meet certain requirements already mentioned above. It must be truthful, with coherence between what is and what is described. It must also be clear, comprehensible, and formative without becoming manipulative. These conditions depend on attention to the semantic question, on rigorous linguistic control and on the distinction between the prescriptive and the descriptive levels. This is not a merely subjective demand but concerns the democratic and constitutional dimension of the communicative and informative process. As public television once had to respect standards of correctness, good faith, and pluralism, so today the new spaces where common sense and consent are formed, such as social media, must respect equivalent duties of accurate information. There exists a close synergy between economic, technological, and media power that strongly influences the formation of public opinion. Language thus becomes a decisive instrument to orient or mislead the behavior of citizens. The control of language through semantic rigor therefore becomes a constitutional and democratic matter, requiring a redefinition of the relationship between traditional and new media. The diversity of media cannot justify the lack of reliable information.

The duties of accuracy already imposed on traditional media must also apply to new media. The absence of such control conceals the values or disvalues underlying misinformation and maintains asymmetries that allow hidden forms of power and control. In this context, it becomes essential not to lose sight of the constitutionalization of the person, the result of a long process

that made the free development of one's personality a fundamental right, a right realizable only if freedom of speech and expression are not instrumentally distorted.

As Frank Pasquale notes, the problem of partisan propaganda has long existed, but an automated public sphere can aggravate it, allowing falsehoods to spread virally. It is necessary to ensure that automated systems respect the same duties imposed on human journalists, to act in good faith and to provide accurate information on matters of public interest, based on verified facts and credible sources. Only under these conditions can individuals understand what happens around them, recognize violations of their rights, and have the tools to act for their protection. The example of data protection is emblematic. The GDPR grants many rights to the data subject, but it is often difficult, for those without expertise, to recognize violations. This difficulty is cultural, social, informational, and linguistic. The right to freedom of information includes a negative right to non-disinformation, deriving from freedom of conscience and thought, the first fundamental liberty of liberalism, which implies the right not to have one's conscience manipulated by false or distorted information. Disinformation on matters of public interest undermines other fundamental rights, such as autonomy and the free development of personality. If, as recognized by the European Court of Human Rights, social media are now essential for exercising freedom of expression and access to information, a corresponding right to non-disinformation must be affirmed. This is even clearer if we accept the definition of the "digital citizen" as one who masters the competences of democratic culture, engages responsibly in civic life, and is committed to defending human rights and dignity.

When information concerns facts or questions of public interest, the highest possible level of semantic control is an indispensable condition for genuine democratic participation.

7. Trustworthy AI: linguistic control as a prerequisite for building a relationship of trust with advanced technologies

In the European context, for more than a decade, since the construction began of the strategic framework needed to govern the development of intelligent technologies, one of the main concepts around which attention has been focused is that of trust. This reference appears in different formulations. We speak, for example, of AI as a force for good in society, of trustworthy AI, of AI4people, and of a good AI society. It is emphasized that precisely because of its wide-ranging impact, AI requires "trust from society" and a social introduction that allows trust and understanding to be built.

In the introduction to the Proposal for the new Regulation on Artificial Intelligence (the Artificial Intelligence Act), the European Commission repeatedly calls attention to "a European approach to excellence and trust", aimed at "developing an ecosystem of trust by proposing a legal framework for reliable AI". It also stresses that the Proposal "aims to give people and other users the confidence to adopt AI-based solutions". The persistent appeal to this concept seeks to characterize the European approach, distinguishing it from others. The idea of trustworthy AI is, in general terms, coherent with the repeated emphasis on the respect for fundamental rights and freedoms of European citizens in the digital transformation of European society. Nevertheless, some critical observations on the use of this notion cannot be avoided.

Criticism of the notion of trust inevitably recalls the objections already raised here to anthropomorphizing ideology and the dialectic of opposites. Although the notion of trust began to be studied as a property of interpersonal relationships based on mutual support and cooperation

only in the 1950s, it is now evident that in common language the term refers to a special bond that arises between the actors in a relationship and implies that these actors are necessarily persons. The notion of trust is the pivot around which human relationships develop. Its meaning involves relational aspects and thus requires the existence of a relationship. It is not the relationship itself but one of its properties. One can think, for instance, of the care relationship between doctor and patient, which is traditionally founded on mutual trust. In this field, the patient trusts the doctor not only for competence and skill but also because of the expectation that personal interests and needs will be respected and met. A relationship based on trust rests on beliefs and feelings, not on certainties, that a person will act in a way that honors the trust placed in them. It is a property of the relationship that projects it into the future, founded, however, on a continuity of behavior that must be confirmed over time.

Applying the notion of trust to technologies can therefore be misleading, since many of these elements are missing or function differently. The element of commitment or disposition to maintain a given behavior is absent, because AI is (pre)programmed to act in that way and does not develop an attitude to maintain it over time, since it cannot act otherwise unless programmed to do so. The projection into the future that relational trust requires is also lacking, because when trust is directed toward technologies, what matters is the contingent action that allows a given problem to be addressed in the present.

If it is true that in ordinary language the notion may refer either to human beings or to technological products (what Giddens calls trust in abstract systems) it must nevertheless be noted that the trust referred to in relation to AI has a metaphorical meaning. Applied to intelligent systems, trust becomes the technical reliability of their performance. Compared to trust between humans, what is missing when the notion is referred to machines is relational reciprocity, as well as the effort required to earn that trust through consistent behavior over time.

What, then, is the problematic aspect of using the term trust in reference to AI? Speaking of trust in artificial systems or technologies means creating "access points" to trust, or rather to that meaning of the term linked to the formation of strong bonds based on beliefs and feelings, which foster an attitude, often unconscious, of humanizing the object to which we attribute trust. Emphasizing the misleading use of this notion in the Proposal for a Regulation does not mean denying that the general European institutional orientation, formally directed toward the protection of fundamental rights, is not shareable. What matters, with this critical remark, is to prevent this legitimate aim from being weakened in other ways, particularly through the use of language.

Another aspect to consider when speaking of trust is the possible confusion between trust and an act of faith. One may say, for example, "to have blind trust in someone". This overlap of meanings is common in ordinary language. Blind trust refers to adherence to what is proposed without questioning it, because certain conditions, such as the presence of an expert, reassure the one who places that trust and lead them to refrain from further verification. In this semantic nuance of trust lies the implicit presence of an authority to rely on, without questioning its precepts and prescriptions. Though subtle, this shift has significant effects if the boundaries of the narrative of AI remain anthropomorphizing and shaped by the dialectic of opposites.

If one follows the narrative of neutral, objective, omniscient AI, one is predisposed to a form of deference toward it, as if it were an authority. There is no need to ask questions or to develop critical thought. The algorithmic oracle tells us everything because it knows everything and owes no explanations for its precepts. If we are inclined to see AI as an authority, the passage from AI that makes predictions about activities and events to one that judges human beings will

be almost imperceptible. It will be difficult to understand that "algorithms are opinions embedded in codes. They are not objective".

If attention is not kept on this point, it will be precisely the notion of trust that undermines European efforts toward human-centered progress. This notion belongs to the conceptual taxonomy of the narrative of AI and, with its persuasive force, risks inhibiting a critical and rational approach to technological development. Indeed, to trust someone means to rely on them, and this attitude often follows from the belief that there is no longer any need for control. But this is another linguistic trap we should avoid.

If too much space is given to the persuasive power of notions such as trust, the role of law itself in guiding technological development risks being emptied. Those who trust do not control, and those who are not controlled do not need rules within which to act.

As will be seen in the following chapters, while it is difficult to keep the discourse on AI always transparent, semantically clear, and balanced, this difficulty does not imply the impossibility of striving to reach that goal to the highest possible degree, especially when legal categories and institutions are at stake. If we wish to be the protagonists of the profound transformation that affects our society, then "it is we citizens who must decide how to distribute the benefits and control the risks linked to the spread of AI. We must ensure that everyone is aware and shares the dangers of monopolistic concentration, inaccuracy, and the pursuit of harmful aims".

Chapter 2

Technological progress and technical problems: the semantics of the *posse*

1. Introduction

In the previous chapter, we examined the concepts and arguments that have shaped, up to today, the narrative of AI addressed to the public and taken up in European institutional contexts. As observed, this narrative developed through a dialectic of opposites nourished by an anthropomorphizing ideology. Its effects concern not only the normative level, with regard to fundamental rights and their underlying values, but also the technical aspects of technological development.

Technological progress risks being misunderstood because of an ambiguous narrative. It is therefore necessary, in this chapter, to focus on the semantics of technical questions. Starting from a brief recall of the main stages in the development of intelligent technologies in the past century, attention will turn to the distinction between technical and ethical problems. This will make it possible to highlight the critical aspects of the idea that speaking of AI means above all dealing with technical decisions reserved for experts.

After clarifying the traps hidden in this conception, the discussion will address the question of knowledge of reality, a theme with a long history in Western philosophy. The aim is to show the problems arising from the widespread belief that through AI, capable of processing immense quantities of data, we can now offer an exhaustive and complete representation of reality. The analysis will reveal that those who uphold the objectivist rhetoric of big data conceal a specific ethical position, according to which technological innovation is neutral with respect to value choices and sociocultural orientations.

2. Brief history of AI

Some authors trace the historical evolution of AI back to the twelfth century, recalling the literature of that period in which a creature called the golem appears, human-like in form but much stronger. Although the origins of this narrative can be so ancient, for the purposes of our discussion it is appropriate to situate the birth of AI in the twentieth century, within the advances made possible by the unprecedented availability of scientific knowledge and its technical conversion. It is in this specific period that human beings, through technology, developed a new power of control over existence. It is precisely in relation to this excess that ethical problems arise concerning the inadequacy of the tools elaborated within traditional ethics, which were mainly, if not exclusively, focused on present action. What intelligent technologies make available to humans is the transformed capacity to act predictively, with consequences for the future, and this makes today's reflection on technology radically different from that of the past.

The term artificial intelligence (AI) was coined in 1956 by John McCarthy, who defined it as the science and engineering of making intelligent machines, especially intelligent computer programs. The path leading to the current developments and applications of AI has not been continuous but marked by so-called winters and springs. These expressions refer to periods of stagnation not only in progress but also in AI research itself, followed by phases of strong acceleration. After several attempts between 1975 and 1985, some successful and others not, investments in AI resumed from 1995 onward. With the advent of the Internet, which required intelligent applications and large volumes of data, AI began to find broad use in many fields and through various tools, such as search engines, website development, and commercial transactions.

Since the 1990s, AI has recorded many successes. It is recalled that in 1997 the computer Deep Blue defeated the world chess champion Garry Kasparov; in 2011 IBM's expert system Watson defeated human contestants in the television quiz Jeopardy!; and since 2015 AI systems have been able to perform oncological analyses in minutes that would take human specialists decades. These successes relied on traditional programs, with manually designed algorithms, capable of performing specific tasks that require specialized expertise. However, the management of complex problems requiring multiple problem-solving skills remained beyond the reach of these technologies.

Things began to change with the development, from 2012 onward, of Deep Learning, a subcategory of Machine Learning. Unlike the latter, characterized by supervised learning, Deep Learning is based on neural networks that use unstructured data in much greater quantities. These networks function similarly to the human brain and attempt to identify, independently of human intervention, the distinctive features needed to solve a problem. It marks the era of self-learning machines.

A first example of unsupervised Deep Learning dates from 2012 with Google Brain, which autonomously identified the concept of a cat after analyzing millions of unlabeled images for several days. In light of this turning point in the evolution of AI, it has been observed that "computers are now educated rather than programmed". The use of the term "educate" evokes an activity traditionally linked to human beings, since for non-human entities such as animals the term "train" is usually used. In the case of AI, however, speaking of "educating machines" has become common, even though the correct term would be "training".

In everyday language, the verb "educate" carries many shades of meaning and refers to a complex process involving intellectual and moral growth, refinement of capacities, and understanding, not merely the execution of programmed tasks. This is yet another example, among many, of how anthropomorphizing ideology is deeply rooted in common language.

Misunderstandings, however, are not induced solely through language. Another important source of confusion, which can mislead even experts, is the lack of distinction between the types of problems that AI raises. This refers specifically to the distinction between technical and ethical problems, a distinction well known in bioethical and environmental ethics reflection. The following section is devoted to this important distinction.

3. Why are technical problems ethical problems?

Since the digital transformation of society became a concrete reality and intelligent technologies can now be applied to almost every human activity, not only their potential and advantages have emerged, but also a number of issues that raise political, social, and ethical questions. It is on these latter that we will focus, as they concern the individual, their rights, and their duties.

In institutional and media debates, the expression "ethical problems" is often used as if its meaning were self-evident. In reality, outside the context of ethical reflection, it is not always clear what is meant by "ethical problems". They can be defined as problems of choice between alternative courses of action likely to produce opposing effects. Such problems may arise in daily life: whether to use medically assisted procreation, to refuse a life-saving treatment, or to donate blood or an organ.

In these situations, those facing ethical choices try to find the best answer based on the values built along their life path. Yet these issues are not only individual but also collective and institutional. It is therefore necessary to ask whether, when they are discussed in institutional or media contexts, we refer to problems whose solution lies within common or shared ethics. As we will see, this is not the case, and the difference between individual choice and collective decision in public ethics is far from marginal.

In institutional contexts, ethical problems are problems of choice that must be addressed through principles and norms, the result of a critical elaboration of common ethics. In other words, they refer to critical ethics, understood as the sphere of human action in which several possible paths are available, and the choice among them relies on normative criteria that invoke moral, legal, and deontological values.

In Europe, these principles and legal norms are expressed through the framework of fundamental rights, which recognizes human dignity as the minimum shared value and the foundation of equality. Unlike in the institutional setting, where the principles of dignity, integrity, freedom, equality, and solidarity guide political decision-makers, in the media discussion of ethical issues raised by scientific and technological innovation the applicable criteria are often unclear. This discourse reflects the ordinary use of the notion of ethics, commonly treated as a synonym for morality. In this context, "morality" is often implicitly taken to mean a correct and objectively grounded morality, rarely seen in its historical and cultural dimension.

As a result, when media communication refers to ethical problems without clarifying the term, there is a high risk that the public assumes these are questions that have objectively moral answers. What is overlooked is that today's ethical questions, whether on controversial existential themes such as the end of life or the beginning of life, or in the field of AI, are matters of public ethics that require the use of normative criteria shaped by the cultural and legal framework of a specific historical moment.

We live in a time in which scientific and technological innovation has radically transformed, and improved, human life. This innovation has also generated an excess of new possibilities for action that demand open-minded analysis. What often happens instead is that the media present ethical debates in polarized form, between two extreme positions. On one side stand the bioconservatives, who see ethical issues as questions to be answered through an objectively grounded morality. For them, technology serves to maintain an existing natural order. On the other side are the transhumanists, who push the use of technology on the human being to the extreme without considering its ethical consequences. Such polarization offers no guidance for governing the entirely new phase humanity has entered.

To govern this phase, normative criteria developed within morality, law, and sectoral ethics are needed. Choices of public ethics are complex processes that, while respecting ethical pluralism, must not lose sight of the minimum value by which public actions are measured: respect for human dignity. This is both a value and a necessary standard for public decisions in European institutions seeking to govern the changing nature of human action and to ensure that technological development remains aligned with the future needs of humanity. Hans Jonas called this the ethics of the future, an ethics capable of reasoning in a forward-looking way to safeguard the interests of present and future generations.

To achieve this, it is first necessary to develop the ability to identify ethical issues, an operation that is neither easy nor obvious. Any attempt to resolve an ethical problem must take into account that ethical issues are often hidden behind technical ones or depend on them. Those who possess expert knowledge, such as computer scientists, often appeal to their technical-scientific mandate and the neutrality of their work to declare themselves outside the scope of ethics. This attitude, common among experts of many technical and scientific fields, has been described in bioethics as the paradox of an ethics that conceals ethics.

Behind the idea of an exclusively technical mandate lies an implicit ethical stance regarding who holds decision-making power and which goals and outcomes are to be pursued. In today's technological world, there is a tacit assumption that the technical solution is the right one and must always be pursued. However, as Garrett Hardin observed, this attitude characterizes human society only since technology became central to economic and social development, and such solutions are not necessarily the best in the long term.

Hardin defined a technical problem as one that requires only changes in the techniques of the natural sciences, demanding little or nothing in terms of changes in human values or moral ideas. A technical problem exists when it can be solved through specialized expertise, institutional strategies, or evolving knowledge that allows us to foresee its imminent resolution. However, the history of technological development shows that seeking solutions only on the technical level limits our understanding of the multidimensional nature of human existence.

If we stop at technical solutions, we risk overlooking the ethical assumptions hidden behind them, which concern decision-making power and the predetermined direction of progress. Some of the ethical assumptions masked by technical solutions are that technology is neither good nor

bad, that it evolves naturally to be used by anyone, that it has an egalitarian vocation, that it needs no ethical or legal regulation beyond self-regulation, that its development is inevitable and thus humanity's destiny lies in its hands, and that data, including our own, serve technology's self-improvement for our benefit.

A major consequence of a predominantly technical-scientific approach to AI is the belief that AI is ethically neutral. In fact, far from being neutral, AI systems are programmed to make choices that reflect specific values. This is evident in autonomous vehicles, programmed to operate according to rules for morally demanding situations, which are decided by the programmer. Thus, while these systems may include ethical rules, they do not understand their ethical quality, and data-based systems would obey unethical rules just as readily as ethical ones.

If one believes that technical issues can always be solved through technical means, it follows that only those with technical expertise can decide. If one also assumes that the solutions to global challenges such as environmental crises are purely technical, there will be no space for alternative paths, since innovation itself will appear as the only way forward. This view excludes the possibility that in some human domains it might be better to proceed more slowly or maintain the status quo, as suggested by Frank Pasquale.

When the search for solutions to human problems stops at the technical level, technicians are seen as the legitimate decision-makers, even when the issues are political or ethical. Those who think that problems such as climate change, poverty, and globalization are purely technical endorse the idea that new technologies and methods are sufficient, assuming that scientific and technological progress will always find adequate solutions. But these solutions often generate new ethical problems due to the greater capacity of modern technologies to affect human and non-human life.

Those who hide behind purely technical solutions perform, as Zuboff writes, an assault on our awareness and at the same time strengthen the ideology of technological determinism. Determinism and neutrality are the key terms of the technocentric vision of society.

This vision is supported by the current value attributed to data in representing reality. Since the twentieth century, the notion of data has lost its character of irreducibility, once considered a limit to knowledge. As Sellars and Goodman observed, there are no data that do not derive from theoretical structures, and thus no ultimate data on which knowledge can be unequivocally founded.

The unprecedented availability of data encourages confidence in the possibility of capturing reality in its entirety. Through data analytics, it is possible to extract probabilistic correlations and base predictions and decisions on them. The notion of data has evolved into that of big data, understood as the ability to act on a large scale to extract new insights or create new forms of value that transform markets, organizations, and relations between citizens and governments.

Big data can permeate social and individual life because of the assumption that everything is representable and solvable through technical means. The technocentric view holds that the sheer volume of data makes them representative of all social and natural reality, promoting a rhetoric of exactitude. Assuming that data are neutral and exhaustive makes their handling a technical issue, but this rhetoric conceals profound ethical questions.

The four V's of big data, that is, volume, velocity, veracity, and variety, are themselves sources of ethical concern. Volume is a matter of perception, historically contextual and dependent

on the available technologies. Veracity cannot be taken as inherent, since data can be manipulated and inherit biases from past collection practices, such as incomplete samples or ideological and racial criteria. Hence, transparency in data collection, management, and control is essential, particularly when the outcomes have large-scale effects.

We are not dealing only with a technical issue of how to ensure transparency but with an ethical one concerning the purposes for which data are used. Transparency, and the related need for explainability in intelligent technologies, has been debated in connection with black-box algorithms, which exhibit three kinds of opacity.

The first concerns intelligibility for users affected by automated decisions, who may not understand the underlying mechanisms. The second involves deliberate opacity introduced by institutions or organizations for various reasons. The third concerns algorithms whose outputs cannot be clearly traced back through their decision-making process, making even programmers unable to explain them.

For the first two, transparency depends on value judgments and balancing interests. The third is the most problematic, since an indecipherable algorithm hinders transparency efforts even in less opaque systems. A possible technical solution is the use of white-box algorithms, whose logic and decision paths are clear, as in decision trees. Yet these are suitable only for less complex predictive tasks, while more complex ones require black-box models, which must later be interpreted through additional analytical methods. This process takes time and precision, and meanwhile, the continued use of such systems raises serious ethical issues.

In justice, algorithmic opacity can undermine constitutionally protected principles such as equality and non-discrimination, as shown by the COMPAS case in the United States. In medicine, diagnostic systems based on black-box algorithms can affect the quality of information provided to patients and, consequently, their autonomy in decision-making.

In summary, technical problems often conceal ethical ones whose resolution requires normative criteria and new categories such as transparency, explicability, and trustworthiness. To conclude, it is worth recalling dataism, a new philosophy or even religion that reduces all human and natural experience to data, assuming that what results from data is certain and exact.

As we will see in the next section, the spread of this data ideology is due to many factors, among which one plays a central role: the failure to distinguish between knowledge derived from the scientific method and knowledge obtained from data analysis by intelligent systems. In other words, the unacknowledged shift from causality to correlation.

4. Knowledge in the technological and in the scientific perspective

The knowledge of possible experience, together with the search for an explanation of what reality is and what its nature is beyond appearance, are constitutive elements of the question of knowledge, one of the oldest themes in the history of philosophical thought. In simple terms, philosophy has always asked whether an objective empirical reality exists, capable of pure description, that is, not distorted by the subjectivity of the observer.

In general terms, knowledge can be understood as "a technique for determining any object, or the availability or possession of such a technique". A technique of determination is any

procedure that makes possible the description, calculation, or verifiable prediction of an object, while "object" refers to any entity, fact, thing, reality, or property that can be subjected to such a procedure. The relationship between the selected object and the cognitive operation concerning it has been interpreted in different ways throughout the history of philosophy. The key terms of this history are transcendentand transcendental. Simplifying their semantic history, up to Kant both terms referred to "the properties that all things have in common, which therefore exceed the diversity of the genera in which things are distributed". From Kant onward, transcendent refers to what exceeds the limits of possible experience, while transcendental refers to what precedes experience a priori yet serves only to make empirical knowledge possible.

So-called metaphysical science rests on an essentialist and transcendent conception of reality. In the previous chapter we discussed essentialism in relation to language, here it is revisited as one of the foundations that long guided scientific knowledge. According to this view, there are ultimate truths that the scientist can definitively establish, and "the best theories, those truly scientific, describe the 'essence' or 'essential nature' of things, the realities that lie beyond appearances". Metaphysics is thus the first science, preceding all others and determining their validity. Its aim is the ultimate and incontrovertible explanation, based on an a priori or speculative approach. This method, assuming fixed truths to be discovered, carries an inevitable dogmatic attitude, a way "to protect and perpetuate achieved values" in the absence of methods for replication and verification such as the principle of falsification introduced by Popper.

Up to a certain point in the history of science, assuming first truths served to guarantee certainty where methods of verification and control had not yet been developed. The essentialist conception was part of Galilean philosophy, which Popper refused to defend. For Popper, who completed the critique of metaphysics, "theories are nets cast to capture what we call the world, to rationalize it, to explain it, to master it". With Popper's theoretical contribution, scientific knowledge assumed its current empirical-experimental character, definitively freeing itself from metaphysical influences.

The criterion identified by Popper to distinguish scientific knowledge from metaphysics is falsifiability. Scientific theories are falsifiable, while metaphysical ones are not. The principle of falsifiability replaced the earlier positivist criterion of verifiability. According to it, a theory is empirical if it can be disproved by experience. Its epistemological advantage is that countless confirmations cannot make a theory certain, while a single refutation can invalidate it.

The overcoming of metaphysics is not only methodological. More broadly, the boundary between modern and premodern culture lies in the radical critique of the metaphysical vision of medieval Christian and Aristotelian thought, which conceived everything in nature and society as having a necessary order and predetermined purpose. In this view, the differentiation of knowledge and techniques was to reflect the order of reality and the essences of things, grasped through real definitions. The turning point of modern science lies in placing inquiry within the limits of experience rather than beyond them. The empirical-experimental method, based on deductive reasoning, best guarantees empirical falsifiability and intersubjective control of scientific theories.

This reasoning differs from that applied by seventeenth-century scientists, who grounded knowledge on observation and, through inductive reasoning, formulated general laws. From Bacon onward, attention was directed to facts and to interaction with the external world. The construction of the empirical-experimental method based on deductive reasoning is therefore relatively recent and was achieved through centuries of refinement by thinkers such as Descartes, who proposed intuition as a step toward hypothesis formation, and Claude Bernard, who distinguished between having an experience and conducting an experiment. The latter distinction marks the boundary

between genuine scientific method, founded on intersubjective control and falsification, and what lies outside it. In this sense, "knowledge becomes a relative process, in which there are no absolute data but only progressive transformation".

This shift enabled the transition from a realist to a constructivist conception of knowledge, allowing us to question absolute claims of scientific objectivity in light of its historical and sociocultural context. As it has been observed, "scientific knowledge does not simply accumulate, nor does technology invariably advance benign human interests. Changes in both occur within social parameters already established". Emphasizing this constructivist view reveals a hidden assumption in the relationship between science and truth. Scientific knowledge indeed allows the pursuit of truth, but since it selects certain areas of experience while excluding others, the truth achieved is limited to those domains and therefore not absolute. It is valid and controllable only within its own context of investigation, as it is repeatable and falsifiable.

The scientific method is based on corroborable hypotheses elaborated by scientists seeking to understand the mechanisms linking events and causes. The study of causality, central to philosophy since Plato, has taken various forms. In its deterministic meaning, causality implies a necessary relationship between cause and event, but this conception was overturned in science by Heisenberg's principle of uncertainty, which abandoned the idea of necessity. In philosophy, the concept evolved from Hume's view of causality as constant conjunction to Suppes's probabilistic interpretation.

The distinctive value of the empirical-experimental approach is not its ability to give ultimate answers, but to provide verifiable and valid ones, answering why something happens rather than merely observing that it happens. To answer why, inquiry must rely on the elaboration of scientific theories and on deductive-causal reasoning.

When we move from these general considerations to the investigation of reality through intelligent technologies, we find that algorithmic procedures underlying predictive models use an inductive method. Such models stop at knowing that something will happen, based on correlation rather than causation. This approach yields results with lower reliability than those obtained through deductive methods.

Anderson has called this passage from seeking reasons to knowing that something happens "the end of theory". No causal hypotheses are needed, only statistical correlations. Paolo Benanti similarly observes that "we witness technological developments (capacity to act) that correspond to no scientific development (capacity to know and explain)". Martin Ebers speaks of a shift "from causation to correlation", noting that most data-mining techniques rely on inductive knowledge and pattern recognition within datasets rather than causal explanation. This correlation-based technique, faster than causal research, fits the economic and social demand for speed. As a result, it is often assumed to be the appropriate model for all socially and economically relevant activities, without questioning whether alternatives are needed, particularly in fields where reliance on predictive models may have serious consequences, such as public decision-making.

As recently noted, relying solely on predictive models in domains like health, justice, or agriculture risks devastating consequences when correlations are mistaken for causation. Confusing correlation with causality leads to the dangerous illusion of certainty, attributing predictive value to what is only statistically relevant. Moreover, inductive reasoning can lead to fallacies of unwarranted generalization, when the data used to train AI systems are insufficient yet still used for projection.

To address these problems, research is moving toward integrating causal reasoning into AI. The emerging field of causal AI has already shown potential in several domains, though practical challenges remain, especially concerning AI's ability to capture causal relations that simulate human reasoning. Many technical factors must be addressed to reduce risks, including data quality, transparency, open-source code analysis, and algorithmic performance assessment.

Moving toward causal AI allows, when used for good, a return to the question of why something happens rather than only that it happens. Reintroducing causal inquiry benefits not only knowledge but also political strategies that aim to keep humans at the center of technological progress. Focusing on causes reinforces awareness of the intentional construction of experience, countering the illusion of an order of things existing without reflection or intent, and prevents reality from being turned into myth, emptying reflective human activity of meaning.

It is crucial not to confuse the modern scientific approach with current forms of knowledge such as machine learning. Machine learning is "a form of numerical pattern finding with predictive power", but its knowledge is inseparable from the computational mechanisms and datasets that produce it. Its outcomes are based on large-scale correlations, not logical causation, and therefore lack the cumulative support of non-falsified hypotheses typical of the physical sciences. In fields like health or justice, the use of falsifiability remains essential to avoid false conclusions. Developing causal AI thus means a return to theory and a step toward overcoming the objectification of individuals, a serious drift in AI use, especially in profiling.

A Council of Europe study on responsibility and AI clarifies that technologies used to personalize products and services tend to objectify individuals, who are no longer seen as moral agents. They are singled out not on the basis of causal theories but of correlations in data, without any reasoned explanation of why they were selected. The systems' logic and processing are highly complex and opaque, designed to extract value from digital traces of behavior rather than to understand its causes. This diffusion of inductive reasoning thus serves those who wish to promote a worldview detached from the search for reasons.

It is precisely the inquiry into causes that opens the space for choice, since explaining why an event occurs allows us to decide whether and how to intervene in the future. This openness to change is inconvenient for those who want individuals to conform to preselected behavioral models. In economic and commercial practices, this has been described as "interested readings of reality". The usefulness of AI systems that merely tell us that something happens, without explaining why, primarily benefits specific social and economic actors. The resulting representation of reality is partial, often biased, and more manipulable than one based on causal-deductive inquiry. To maintain greater control over our understanding of reality, it is therefore essential to develop intelligent technologies capable of providing reasons for the results they produce.

Although causal AI carries its own risks, especially the danger of deterministic interpretations of human behavior, such as reducing actions to physical, genetic, or neurological causes, it remains the approach closest to knowledge that deepens understanding by explaining why events occur. Such knowledge opens individual and collective horizons of choice, laying the groundwork for social programs that protect justice and dignity. This is particularly urgent in a time when it is necessary to overcome self-interest and act so that the consequences of our actions "remain compatible with the permanence of authentic human life on Earth".

5. The technical problems and the semantics of anthropomorphization: domination and power over the individual

Summarizing what was observed in the previous section, we can borrow Frank Pasquale's words: "those who deal with ethics and technology are increasingly realizing that the main problems of AI do not concern technical aspects but its social role". This role unfolds along different, yet interconnected, directions. To identify them, we can draw from Michel Foucault's four fundamental types of technologies. First, AI impacts the technologies of production, aimed at creating and transforming objects. Second, it has taken over the technologies of language, which concern the use of signs, meanings, and symbols. Third, intelligent technologies are technologies of power. Finally, AI is also used in what Foucault calls technologies of the self, which allow individuals "to carry out, by their own means or with the help of others, a certain number of operations on their own bodies and souls".

In this sense moves the debate on human enhancement, which will be addressed in Chapter Four. For now, it is enough to note that the debate concerns the use of pharmacological and non-pharmacological means, such as neurotechnologies (for example, Brain-Computer Interfaces, BCIs), to enhance or improve the cognitive and physical abilities of healthy individuals. The social effects of neurotechnologies, which rely on AI, are both positive and negative. The positive effects relate to improving quality of life through a better understanding of the potential of the human brain. The negative ones concern the abuse of these technologies to control and manipulate the brain externally. We may face a paradox: our brain could be controlled by an AI that acts sine intelligere, without understanding, yet in the hands of those who seek control over others, it becomes a powerful instrument for redefining social order. As Laurent Alexandre has written, "the remarkable progress made recently in reading the brain may lead us to fear the dissolution of our individuality into a vast global 'hub' of consciousness. There is a risk that political powers will show a particular inclination to expand techniques of thought transmission. The price to pay for such security would, of course, be absolute political control. A perfect democracy, in which every citizen is in constant communication with the minds of their leaders, would also be the death of democracy".

Ultimately, underlying the control over the brain lies control over the narrative of autonomy, freedom, and (un)predictability, that is, over those notions that constitute the conceptual taxonomy of the anthropomorphic AI narrative by excess, as discussed in the first chapter.

6. The technical problem that conceals a sociocultural problem

Many voices have raised concern over the growing belief that political, social, criminal, economic, and occupational decisions can be justified simply because "the computer or the algorithm said so". There is an emerging suspicion that it may no longer be humans who make decisions but machines and intelligent systems. In short, we are forgetting that "the world, besides being shaped by natural evolution, is actively created by the cultural development of human beings, with the support of technology". This happens for several reasons, but the constant emphasis on technical aspects, or rather on the purely technical mandate of ongoing transformations, plays a central role in obscuring the human component. To explain this, one can refer to the notion of *Kulturtechniken*, translated into English as cultural techniques. This term designates "operative chains composed of actors and technological objects that produce cultural orders and constructs which are then installed as the basis of these operations. At the core of cultural techniques is the notion that fairly simple operations coalesce into complex entities, later viewed as the agents or

sources running these operations". In this way, procedural chains and linking techniques produce notions and objects that carry essentialized identities. Following this perspective helps explain how the belief arises that algorithms today act as autonomous decision-makers, whether used in justice, medicine, or other areas. Thinking in terms of cultural techniques rather than technology highlights the constructed nature of reality, with a strong subjective component that allows rejecting the dogmas of neutrality and objectivity. It also shows that this construction relies on distinctions derived from data analysis and statistical samples, which are never complete or representative of the whole of reality. This analysis depends on the meanings with which the AI system has been trained, meanings that are historically determined, since AI predictions necessarily rely on past data requiring delimitation and semantic definition. Because the intentional construction of reality and experience "in every kind of anticipation, whether artistic, utilitarian, technological, social, or moral" risks otherwise failing to achieve its intended purpose, it becomes clear why it is necessary to overcome algorithmic opacity, as far as possible, and to ensure transparency. This does not mean achieving total transparency at all levels, but rather providing information that is useful in a given situation, so that users, whether producers or end consumers, have the means to claim responsibility when problems or harm occur. In this direction, Margot Kaminski's proposal to create a taxonomy of transparency for accountable AI is significant. Unlike many authors who propose different forms and degrees of transparency, she distinguishes two main kinds: individualized transparency, which concerns informing and protecting individuals and making the system's logic traceable and justifiable, and systemic transparency, aimed at revealing errors, biases, and discrimination both in the machine system and in the human one that should oversee it, so that corrective measures can be taken. Adapting transparency to the final recipient of information can have the advantage of avoiding an overload of unnecessary data that the user may struggle to understand, but it should not become a way to bypass information duties or maintain technocentric attitudes. Algorithmic transparency depends on both procedural and technical properties of the system and therefore presents significant challenges. In any case, transparency's main goal is to enable responsibility, in the sense of accountability. This direction was already indicated by the Panel for the Future of Science and Technology (STOA) in 2019, which emphasized that transparency is primarily a technical issue, while accountability is an ethical and legal one. It is therefore necessary to move deeper into legal questions to understand how technical issues can be addressed by law.

Chapter 3:

Scientific and technological progress and the legal problems: the semantics of *licere*

1. Introduction

In the previous chapters, we examined the complexity of the semantic narrative of AI, which unfolds between a deep-rooted dialectic of opposites and an anthropomorphizing ideology. We also recalled the problems arising from deterministic conceptions of technological development. In addressing the linguistic aspects that enable the instrumental use of narratives about so-called intelligent technologies, we focused on linguistic traps. On the technical level, we revealed the ethical positions underlying the idea that decisions concerning the path of technological innovation belong exclusively to the technical-scientific domain.

In this chapter, continuing ideally from the previous ones, we turn to the legal articulation of the narrative on AI. The aim is to bring out the implications of the dialectic of opposites, of

anthropomorphizing ideology, and of objectivist rhetoric for the legal dimension of this phenomenon. To proceed in this direction, it is necessary first to recall the complex relationship between law and technology. We will then focus on aspects more directly related to the legal regulation of AI. In order to identify the legally problematic profiles of AI's ideological narrative, we will consider the categories most exposed to the influence of this dialectic of opposites, given the ubiquitous use, in both ordinary and legal language, of the notions expressed through such categories. Examples include the category of legal personality and the related concept of the person.

Finally, we will reflect on the adequacy of fundamental rights as both a measure and a limit to what technological innovation allows us to do.

2. What relationship is possible between law, science, and technology?

In current European institutional discussions, one of the central topics concerns the relationship between technology and law. The question being addressed is formulated as follows: What if law shaped technology? What if technologies shaped the law? These issues are discussed in a series of articles published in the What if collection of the Panel for the Future of Science and Technology (STOA) of the European Parliament, dedicated to these themes. The titles themselves present the problem as an alternative, simplifying the complex interactions between law and technology. In reality, the question is to what extent technology determines the direction law should take and, conversely, to what extent technology can be regulated by law. Similar questions have also arisen regarding the relationship between science and law.

As is well known, the history of law is largely built around identifying the features that distinguish what is legal from what is not. This inquiry has been central to analytic-linguistic legal philosophy. Practiced in this perspective, it has helped to clarify, on one hand, what differentiates law from other normative systems, and on the other, to provide the foundations for rationally addressing the so-called is-ought question.

In other words, this clarification has helped to identify epistemic limits within the normative sphere, those that "differentiate the contours of law from other normative systems with which it coexists in the social space". The question of the boundaries of the legal has long been at the center of positivist efforts to separate law from morality, ensuring the autonomy of positive law from natural law. Yet this same philosophical approach also helped to delimit the relationship between law and nature, between ought and is, distinguishing two discursive levels: one assertive-descriptive and the other prescriptive-evaluative. This distinction between language functions forms the basis for understanding the relationship between science and law. Science expresses itself through assertions, which can be verified as true or false, while law expresses itself through prescriptions, which are neither true nor false but aim to guide behavior toward desired outcomes.

This linguistic distinction reveals the differing methods and purposes of science and law. Science formulates hypotheses about events, phenomena, and behaviors, confirming them through intersubjective verification and explaining them according to causal principles. Law, by contrast, regulates human conduct through the principle of imputation. Legal rules specify the principles and values deemed relevant in a given historical moment to ensure social order and peaceful coexistence. Given the methodological and teleological differences between law and science, a key issue arises: whether, and to what extent, science can be regulated by law and, at the same time, what contribution it can offer to law.

To answer the first question, an important clarification is needed. Asking whether science can be regulated by law does not mean questioning the procedures or experimental methods underlying scientific progress. The issue is never an evaluation of the experimental method itself. What can be subject to ethical and legal assessment are the purposes and uses of the results achieved by that method. Such evaluation is not only possible but necessary if one rejects objectivist and deterministic logics, recognizing that scientific knowledge exists within the social and cultural context of its creators.

In short, decisions about which research to encourage and which results to apply necessarily involve ethical, social, cultural, and legal evaluations alongside scientific ones. Ethical assessment of the sustainability of scientific innovation and the resulting legal regulation become essential because each step forward in science opens new horizons of choice. This excess, born from new knowledge about natural mechanisms, marks the gap between what has been done and what should be done next. Science tells us how something is, while on the ethical-normative plane we decide how that knowledge should be directed toward certain goals rather than others. Scientific knowledge influences normative systems indirectly but profoundly, continually raising questions about the values that should guide the use of results and technical applications to improve individual and collective life.

These evaluations are inspired by science, as its results help eliminate conditions shaped by prejudice, for instance, by clarifying biological processes and dispelling misconceptions, but they do not derive directly from scientific data. To be useful for social, cultural, or criminal policy, they require further deliberation on the values guiding choices within a specific political and legal framework. Those who hold this view argue, on one hand, for the possibility of a rational ethical-legal discourse and, on the other, against the reduction of value to fact, supporting instead an ethics not ontologically grounded but built upon the responsible choices of individuals and communities.

The considerations on the relationship between science and law prepare the ground for addressing the connection between technology and law, which is the focus of our inquiry. In European institutional debates, this relationship is often framed as an alternative between technology as an object of regulation and technology as a regulatory agent. We will first address the former and then turn to the latter in the following section.

2.1. Technology as an object of regulation

Addressing the issue of technology as an object of regulation from the European institutional perspective means focusing on two main questions: whether and how law can regulate the products of technological innovation. Two types of regulatory intervention are usually identified. The first is reactive or ex post facto, while the second can be described as ex ante facto, preventive or anticipatory. The reactive approach represents the traditional way law intervenes: it seeks within existing norms a response to guide behaviors that have already produced effects but that may constitute new situations not directly covered by preexisting rules.

Normally, it is through judicial interpretation that new situations, not yet explicitly foreseen by existing norms, are legally defined. In the technological era, this reactive approach is often criticized as inadequate, since legal intervention takes place only after harm has occurred and is seen as incapable of keeping up with the exponential acceleration of technological development.

To address these limitations, efforts have been made to identify legal tools that can keep pace with the rapid evolution of technology. The goal has been to find instruments capable of preventing potential harms or complex conflicts of interest. Among these instruments are sunset clauses, which allow laws to be reviewed after a few years to assess the need for revision; horizon scanning, which anticipates risks and opportunities; innovation deals, designed to support innovative enterprises and prevent regulatory burdens from hindering their activity; and regulatory sandboxes, which enable companies to test products in a controlled, real-market environment with fewer legal constraints.

From this overview of preventive regulatory techniques, it emerges that law offers several alternatives capable of ensuring damage prevention, risk prediction, and the protection of fundamental rights of users of such technologies. This clarification is necessary, since the regulation of technological developments is often accompanied by an implicit criticism regarding the law's real capacity to keep pace with technological acceleration, the so-called pacing problem. This criticism, typical of certain philosophical currents but also echoed by some jurists, risks playing into the hands of those who wish to free technological development from any form of regulatory constraint. Those who take this stance emphasize the supposed inadequacy of law to regulate scientific and technological progress because of its inherent inability to govern what is constantly evolving.

Within the group opposing legal intervention, two main tendencies can be distinguished. The first is that of those who believe that corporate self-regulation is sufficient to ensure the proper use of technology. Large multinational corporations, often aligned with transhumanist ideology, take this view. This philosophical position advocates allowing the free market to determine the rules through supply and demand, viewing law as an obstacle to this model. From this perspective, broad freedom of action is demanded for companies and research investors, resulting in wide grey areas, activities not necessarily unlawful but largely beyond the control of national and supranational law. Unlike bioconservatives, who, through a strongly ideological use of law, seek to impose excessive constraints on technological innovation, transhumanists aim to stay distant from legal solutions, at least until economic and power interests in a given sector have become consolidated. Among the types of legal interventions mentioned earlier, regulatory sandboxes are particularly encouraged in this view, as they allow wide flexibility before definitive rules are established.

If these are some of the main challenges in the relationship between technology and law, an additional one arises from the specific nature of the technologies in question, those considered intelligent. Up to this point, technology has been examined as an object of regulation, with law retaining its traditional role and technological trajectories moving within the established framework of legal order. However, the emergence of intelligent technologies capable of directly influencing the achievement of legal objectives has brought forth a new approach. In this new view, technology is not merely an object of regulation but also functions as a regulatory agent, comparable to law itself.

2.2. Technology as a regulatory agent

The paradigm shift expressed through the notion of "technology as a regulatory agent" is both significant and symbolic. Until the advent of intelligent technologies, the most effective instrument for guiding and controlling human conduct had been law, thanks to its coercive force. In the age of AI, the question arises as to whether this still holds true.

The new technologies make it possible to act directly on individual conduct, immediately realizing the intended effect. The applications that allow remote control of tools and devices, such as cars, make it possible to act directly on the object, rendering it unusable by the user whose behavior is to be influenced. For example, a driver who has not fulfilled insurance obligations may be prevented from starting the vehicle. Another example is the limitation on the number of copies of a digital work. In such situations, the technological instrument allows one to act directly to obtain the desired result. In other words, technology directly determines what can and cannot be done.

It is necessary, however, to clarify the meaning of "to be able". With such instruments, the possibility of disobeying norms is effectively eliminated, since while legal norms "merely establish what people must or must not do, technological artifacts have the capacity to determine directly what people can or cannot do", making superfluous the intervention of an authority that would otherwise punish those who violate the law. The technology renders the contrary action impossible in fact, making immediately executable what in the legal domain requires the mediation of a judge.

The difference between the norms of conduct of a traditional legal system and one that employs technologies capable of direct control lies in the way "duty" and "power" are conceived. Whereas legal norms traditionally establish duties whose violation requires the intervention of a third party to impose coercive measures, intelligent technologies eliminate this passage, rendering noncompliance de facto impossible. The use of such technologies can offer advantages in some areas, such as contractual and insurance relations, where the technological inhibition of certain behaviors may guarantee the certainty of execution of contractual obligations.

Nevertheless, some problematic aspects must be considered. It is necessary to avoid that these technologies, capable of acting directly on conduct, create asymmetries between contracting parties or manipulate the individual's power of choice, especially through technological interventions that act directly on the brain. Neurotechnologies already allow manipulation of emotions, memories, and sensations for therapeutic purposes, and it cannot be excluded that such technologies might one day be used to condition the behavior of healthy subjects.

Even if technology can act directly on conduct, it does not follow that it is a regulatory agent in the same sense as law. The expression is misleading because it exploits the widespread anthropomorphizing ideology that falsely identifies technology with organs responsible for lawmaking or law application. This confusion equates the coercive modalities that technology makes immediately operative with the authority to decide which behaviors to regulate and in what way. If this distinction is maintained, it becomes evident that technology is not a regulatory agent but a possible means adopted in specific contexts to achieve juridical results. In other words, it is a political decision to allow contracts and laws to be written directly in computer code and automatically executed under predetermined conditions, or to promote compliant behavior through a "techno-nudge" managed by software. From this perspective, technology is a means of executing norms, not an autonomous regulatory agent.

This delegation must be explicitly and clearly defined to avoid misunderstandings regarding the roles of the actors involved. Since it is a choice operating on the institutional and political level, it must remain within the framework of fundamental rights. This has been the European approach in the field of data protection, where an effort has been made to maintain a balance between the use of automated decision-making systems and the right to understand and contest such decisions, including the right to human intervention when automated decisions concern sensitive aspects of individual life.

At the foundation of data protection law lies the goal of maintaining strong juridical control over technological power. This control prevents distortions in the application of computational law and preserves areas where it does not apply. It is worth asking whether it is appropriate to preserve contexts in which compliance with a juridical norm follows its traditional sequence: duty, noncompliance, intervention by an authority. This regulatory modality, based on the awareness of consequences and on the freedom to choose whether to comply or not, corresponds to what Herbert Hart defined as a choosing system. It presupposes the individual's freedom and responsibility.

If juridical norms are implemented through technologies that inhibit noncompliance immediately, making contrary action impossible, the juridical model inspired by law as a system of choice is reduced or even surpassed, especially when such technologies extend to the majority of juridically regulated activities. While mechanisms that prevent noncompliance may seem advantageous in human relations, especially in contractual contexts, in penal matters the reflection cannot stop at mere practical utility.

In criminal law, the issue of freedom of choice as the foundation of penal action remains central. The penal system has historically evolved from objective liability to the affirmation of personal responsibility linked to will. The postulate of modern criminal law is the freedom of the individual to act or not in accordance with norms and the consequent personal responsibility. This principle grounds the possibility of re-educative and rehabilitative justice. Extending intelligent technologies to areas where fundamental social values are at stake risks shifting criminal law from a system of choice to a system of social defense, transforming the conception of the person from a free and responsible agent into one determined by external forces.

A penal law oriented exclusively toward social defense presupposes an ethics without freedom. The adoption of intelligent technologies in sensitive contexts must therefore be evaluated by democratic institutions and not left to hidden powers that protect only their own interests. It is for parliaments to decide when and how technology may be used coercively. Although technological innovation is a valuable resource for the law, it is essential to prevent abuses that could threaten personal liberty.

In conclusion, far from being a regulatory agent like law, technology is a means at the service of juridical objectives. Preserving the distinction between the role of law and that of technology prevents improper incursions into the juridical sphere by economic and technological powers and allows for a relationship between law and technology that is coherent and functional for both.

3. Legal categories tested by AI

The reflections developed in the previous paragraphs aimed to show the complexity of the relationship among science, technology, and law, especially regarding the clear definition of their respective roles in determining the directions of technological innovation and identifying those responsible for such choices. Decisions about the possible uses of intelligent technologies cannot and must not belong solely to technicians, computer scientists, or actors with specific economic interests. These decisions belong to political authorities and those responsible for lawmaking. It is their task to establish priorities and to assess which uses are compatible with socially and culturally shared values. For these decisions to be sustainable and coherent with the underlying value framework, policymakers must resist the influence of transient fashions and ideologies rooted in common sense and irrational dialectics. I refer here to the anthropomorphizing ideology and its

related dialectic of opposites. At certain moments in European institutional history, these tendencies entered political debate and even threatened to invade law and its categories. One notable case was the 2017 debate around the European Parliament Resolution with recommendations to the European Commission concerning civil law rules on robotics. The Resolution received wide media attention for its proposal to attribute electronic personality to robots. The idea was to extend legal personality to robots with certain characteristics: autonomy, self-learning capacity, and unpredictability of actions. The proposal sparked intense discussion, leading many legal experts to sign an open letter urging the Commission not to proceed. In its 2018 official response, the Commission avoided the issue, which is now considered surpassed in the European context. Yet in other parts of the world, controversial choices have been made, such as granting citizenship to a robot.

Years later, scholars often describe that initiative as a case of sci-fi speculation, relieved that the current European draft regulation no longer contains such elements. Still, it is useful to revisit that debate for two reasons. First, it reveals how deeply rooted the anthropomorphic vision of technology has become, reaching the core of legal categories such as personhood. Second, it allows us to understand the legal reasons why this category cannot be extended to intelligent artifacts. Historically, the relationship between jurists and legal categories rested on a natural law conception that denied the historical dimension of law and promoted a static, immutable view of its categories. Law understood as natural law was meant to preserve the enduring values of the individual. In this view, legal categories safeguard a preordained order of things centered on the human being. These orientations were influenced by an essentialist vision of reality that produced a unified and unchanging image of legal concepts and categories.

Even apart from natural law, the rigidity of legal categories derives from the conceptualist attitude of legal doctrine, with roots in the French School of Exegesis and the German School of Jurisprudence of Concepts. Though these orientations faced deep criticism, their assumptions persist in the attitudes of many modern bioconservative jurists who reject reconsidering legal categories as a legitimate response to new and unprecedented situations. For them, legal categories must not be extended or interpreted differently from how they have been so far, and on this basis they reject extending legal personality to robots.

This position, however, is not rationally sustainable. The analytical-linguistic philosophical perspective challenged the idea of fixed legal categories. Seeing law as language makes evident that person, responsibility, and capacity are normative concepts whose meaning can be modulated through definitions. Extending legal personality, capacity, or responsibility involves determining the class of entities referred to by a term (denotation) and the properties they must share to belong to that class (connotation). From a legal-linguistic standpoint, this is an interpretative choice that may be more or less inclusive. Yet, because it occurs within legal language, the choice is not arbitrary but depends on considerations of opportunity and axiological coherence with the values of the legal system. Extending the meaning of person to include robots is not a neutral operation free of symbolic and value implications. Determining who counts as a legal person entails balancing values, which can narrow or broaden the set of entities entitled to legal protection. This explains historical differences among legal systems regarding who is recognized as a person: slaves were not, unborn children have been treated differently across contexts, and some legal systems now recognize nature as a person. Extending the semantic scope of person brings significant benefits to included entities. To be a legal person means to enjoy the highest protection, being the bearer of fundamental rights. When this extension concerns non-human entities, it implies not only the attribution of rights and duties but also an evocative symbolic equivalence between those entities and human beings.

The question of granting legal personality to robots thus involves rights traditionally reserved for humans. From a technical-legal standpoint, no insurmountable obstacles exist, since person is a normative concept adaptable to the purposes of law. The real question is not whether the operation is technically feasible but whether it is ethically sustainable. Extending legal personality to robots requires addressing: (1) the reasons motivating this choice, (2) who is authorized to make it, and (3) the implications for the constitutional notion of personhood. Through constitutionalization, person has lost its purely technical neutrality and has become a legal notion functional to constructing personal identity and expressing the system's foundational values. Within this framework, extending the category to include intelligent robots has a symbolic and axiological significance that goes beyond finding a legal regime for liability issues. It reveals how deeply anthropomorphizing ideology has penetrated institutional thinking, even seeking juridical legitimization. Although robots with AI possess particular characteristics, they remain devices that require a specific and detailed robotics regulation, based on the premise that what must be regulated are not the robots themselves but the behaviors of people who design, build, market, and use them.

Recalling the institutional episode of e-personality is not meant to reject technological innovation but to highlight the legal instrumentalization behind such proposals. At their core lies the attempt to shift responsibility for damages caused by complex systems across multiple actors: creators, programmers, producers, and users.

The issue of liability is now addressed in the recent European proposal for an AI Regulation. Unlike the previous Resolution, the proposal focuses on the level of risk of AI systems and adjusts liability according to that risk and the role of the actors involved. It assigns differentiated obligations to providers, manufacturers, importers, distributors, and users, proportionate to their function in the process, from design to market introduction, and to the level of risk generated by AI systems.

This risk-based approach distinguishes among unacceptable, high, and low-risk AI systems. The first are prohibited, the second permitted under strict conformity assessment, and the third subject to transparency and information duties. The proportional approach aims to prevent exclusive reliance on self-regulation by private entities. While self-regulation can support compliance through incentives, it must be monitored to avoid becoming a façade serving private interests. In past debates on corporate criminal liability, self-regulatory codes often failed to reflect real corporate culture, functioning instead as a superficial display of compliance.

Within the same European proposal, attention should be paid to expressions such as "ecosystem of trust" and "trustworthy AI", and to the decision to frame the user's relationship with AI products in terms of trust. This focus, already noted in the first chapter, is misleading though useful for marketing. It shifts attention away from ethics, reducing the user to a passive consumer who must rely on the technical assurances of safety. Trust becomes founded on technical compliance alone, shaping public perception and advertising narratives of AI.

This impression is confirmed by Thomas Metzinger, a member of the expert group that produced the 2019 Ethics Guidelines for Trustworthy AI, who noted that the narrative of trustworthy AI serves to develop future markets, using ethics as a decorative frame for large-scale investment strategies.

The semantic and political shift from ethics to trust carries risks. The notion of trust tends to obscure decision-making autonomy, emphasizing instead reliance. Though trust here is grounded in technical guarantees, it is almost blind, since the average user cannot evaluate or

question the reliability of the product. Users are asked to trust without clarity about whom that trust is placed in: the AI, the programmers, the producers, or the distributors.

This reliance-based strategy is faster and cheaper than promoting serious user education, understood not only as technical training but as broader critical literacy.

In conclusion, for the narrative of trustworthy AI, which in the European vision should sustain a human-centric approach to technological progress, to be more than rhetorical, the construction of the ecosystem of trust must begin with genuine user education. When founded on individual autonomy, trust ceases to be blind and becomes a dynamic process of building a relationship with the technological, never entirely complete, where the human person remains the measure of trust.

4. Are fundamental rights an adequate framework to regulate technological innovation? The case of facial recognition

The question posed in the title invites reflection on the adequacy of the current European framework of fundamental rights in guiding technological innovation. The pervasive nature of such innovation requires careful consideration of its risks and benefits, and fundamental rights serve as the measure of this balance.

Appeal to rights is today the political strategy adopted by European institutions in regulating AI. The Proposal for a European Regulation itself affirms the goal of ensuring that European citizens benefit from new technologies developed and operating in conformity with the values, fundamental rights, and principles of the Union.

This raises the long-standing issue of the effectiveness of fundamental rights. Norberto Bobbio, in L'età dei diritti (1996), noted that the central problem is not to identify new rights but to guarantee the effective protection of existing ones. Having rights on paper is not enough; they must be realized. Yet this remains the most difficult aspect to achieve, even within the cradle of fundamental rights. The tension between these rights, understood as a strategy for protecting the weaker party in any legal relation, and the constant aspiration of those holding power, whether political, economic, or technological, to act legibus solutus is still unresolved. Strong powers adapt poorly to the limits imposed by fundamental rights and often seek new ways to circumvent or weaken them. Their strategy usually does not consist in openly rejecting rights but in adopting hypocritical or compromising attitudes toward them, for example through ambiguous formulas in legislative drafting.

Such formulas are not technically incoherent from a legal standpoint but are axiologically incoherent with the value system expressed by the framework of rights. A clear example appears in the current European Proposal on Artificial Intelligence regarding facial recognition. As previously observed, the proposal is based on a risk approach. It defines a hierarchical classification of technologies by decreasing (in)acceptability: from unacceptable risk, which entails prohibition of certain AI practices (Article 5), to high-risk systems (Article 6 and following), to low-risk systems (Article 52), which are subject only to specific transparency obligations.

Facial recognition is included in the proposal under the category of "real-time remote biometric identification systems in publicly accessible spaces". These systems detect, compare, and identify biometric data without significant delay in public places, regardless of access conditions.

Real-time remote biometric identification in public spaces is listed among AI practices prohibited by Article 5. However, this is not an absolute ban like that imposed on social scoring. It applies only to law enforcement, with exceptions allowed for specific purposes: (1) the targeted search for potential victims of crime, including missing minors; (2) the prevention of specific, substantial, and imminent threats to life, physical safety, or terrorist attacks; and (3) the detection, localization, identification, or prosecution of offenders or suspects covered by the European Arrest Warrant Framework Decision.

A first critical aspect is that this prohibition concerns only law enforcement, while private entities meeting certain conditions defined in the proposal could use such systems. This creates a risk of mass facial recognition. Even though control mechanisms such as conformity assessment are foreseen, ensuring real protection of the rights endangered by such AI practices will be difficult. Privacy is the central issue, but its violation can trigger further discrimination, often affecting the most vulnerable groups. What is at stake are personal freedom and the right to the free development of one's personality, which are undermined when anonymous spaces of movement disappear. Widespread facial recognition may encourage behavioral conformity, not as conscious adherence to norms but as homogenization that erases individual diversity of expression.

A second problem concerns the possible emptying of rights and their function of empowering ordinary people. When the Proposal was presented, several actors, such as politicians, members of the expert group who drafted the 2019 Ethics Guidelines for Trustworthy AI, and civil society representatives, openly opposed the compromise on facial recognition included in the text. Their criticisms also appear in the European Parliamentary Research Service's analysis on facial recognition. At the national level, the Italian Data Protection Authority, in line with the Council of Europe, expressed opposition to the use of the SARI Real Time facial recognition system in its 25 March 2021 opinion. Among the main objections to indiscriminate facial recognition is that it inhibits the free development of personality.

This right is central to European constitutionalism. It allowed the transition from the abstract notion of the legal subject to the concrete person, whose individuality must find realization consistent with their specific nature. In this sense, the individual becomes the "builder of their own personality". For such self-realization to have space in a technologically permeated society, one must ask whether, in certain cases, firm boundaries should be drawn, for instance, a prohibition on facial recognition. Without a decisive stance, compromises that are unsustainable in value terms risk undermining the creation of the very "ecosystem of trust" that the Proposal claims to pursue.

The framework of fundamental rights enables self-realization only if taken seriously and carried through to completion. Its adequacy to keep pace with the profound transformations of the human brought by science and technology depends primarily on the seriousness with which society addresses the question of which values should prevail when balancing interests. The principle of the free development of one's personality remains the guiding star for realizing the constant call to keep the human being at the center. Yet this expression, often used rhetorically, must be made effective to have real meaning.

4.1. The Precautionary principle and the effective protection of rights

With regard to intelligent technologies, maintaining the centrality of the precautionary principle serves the goal of individual self-realization. Conceived to address the profound

uncertainties projected into the future of scientific and technological development, this principle has long guided European environmental policies but extends far beyond that domain. It informs decision-making where risk is high and largely unknown. Broadly, it expresses a call for caution and foresight when decisions involve activities with potentially negative and unpredictable consequences. In a stricter sense, it allows for regulatory intervention, even prohibitive if necessary, despite the absence of definitive proof of harm. It reconciles different interests while protecting those in the weakest position and can therefore be considered "a method by which to guide the necessary interventions of a legal rule capable of preserving reference to fundamental rights".

This principle is more incisive than the principle of prevention, differing from it by allowing action not only in response to proven risks but also in situations of excessive uncertainty, to anticipate strong protective measures for fundamental rights. The precautionary principle provides valuable guidance for decisions on technological innovation, provided it is not used instrumentally to defend ideologically anti-scientific positions. It is not opposed to science or technology but to the abuse of scientific-technological power and its growing excess. It thus acts as a barrier against strategies that empty fundamental rights of meaning. For this reason, it is regrettable that the European Proposal on Artificial Intelligence makes no explicit mention of it. Reflection on this omission is necessary, given the moral significance of the principle.

Its moral value lies in recognizing that what human beings create can produce unforeseeable risks, implying a humble acknowledgment of the defectiveness inherent in human creativity. Recognizing this defectiveness does not mean condemning it but finding ways to protect the human and to make such imperfection a criterion for the governance of public life. In this light, it becomes easier to counter the naive yet insidious anthropomorphism of the mechanical that this volume has discussed.

A further task for policymakers who take fundamental rights seriously is to uphold and implement the principle of explicability.

4.2. The principle of explicability and rights brought to fulfilment

This principle was explicitly formulated in 2018 in the document AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. The document adopts from bioethics four guiding principles for technological innovation: beneficence, non-maleficence, justice, and autonomy. To these it adds a fifth, explicability, considered the means of enabling the others through intelligibility and accountability.

Explicability addresses two issues: a technical one, concerning the possibility of understanding how an artificial system produces certain outputs, and an ethical-legal one, concerning responsibility and the identification of the actor accountable for the (mal)functioning of the algorithm. It relates to the technical problem of opacity in some algorithmic systems, the so-called black boxes, already discussed in the previous chapter. Over time, this principle has been widely taken up in both scholarly literature and European institutional documents, particularly through the related notion of transparency.

The aim here is not to retrace the technical difficulties of achieving full explicability at the current stage of development, which may find new solutions over time, nor to analyze all related concepts. Rather, the purpose is to argue for maintaining this principle, within the limits of its

technical feasibility, as a foundation for human-centric AI and for realizing the right to the free development of one's personality.

There are human activities where adherence to this principle and the technical effort to ensure explicability and comprehensibility of intelligent systems, to the highest possible degree, should be shared by all who care about humanity's future. Among them, justice and medicine are central to European institutional debate. In medicine, characterized by trust between doctor and patient, and in justice, marked by the delicate balance of interests, especially in criminal law, the problem of explicability arises both when algorithms filter data and when they externalize decision-making from humans to machines or software.

In institutional discussions, some suggest that explicability and comprehensibility will gradually be abandoned as algorithmic systems become standard and their use gains the aura of reliability. In such a scenario, professionals – judges or doctors – would rarely depart from system outputs. As using GPS has replaced paper maps, decision-making too may be delegated entirely to AI, without questioning whether the proposed solution is truly the best. As observed, "whenever a decision-making automatism is introduced into a deliberative process, it tends to capture the decision or make it extremely difficult to disregard".

This outcome should be avoided, especially in justice and medicine, for two reasons. First, respecting professional expertise requires that specialists, even without obligation, retain the possibility to understand the systems on which they rely. Mastery and knowledge of AI sustain the professional's self-realization through their work, which demands critical spirit and attention to those entrusting them with vital decisions. Avoiding blind reliance also prevents the human decision-maker from being relieved "of the burden and risk of having to justify and answer for their reasoning". Contrary to views favoring the substitution of human work with technology, the willingness to retain control and responsibility may help preserve professions such as radiologist or orthopedist that risk disappearing. The disadvantages of losing these and other forms of expertise, as well as the capacity of AI to create decent jobs in exchange, remain unclear.

The second reason concerns those subject to AI-based decisions – medical or judicial. Denying them access to relevant aspects of the process that led an artificial system to decide their fate places them in total subordination, not to the human decision-maker but to the machine. This not only undermines trust in systems beyond human control but leads to dehumanizing outcomes, as those affected by such unchallengeable decisions face constant risk of discrimination.

Maintaining explicability is also important because, consistent with the precautionary principle, at the beginning of such a vast transformation of human activities, we cannot yet predict the risks of fully surrendering decision-making to machines.

The specific role this principle can play in medicine will be examined in the next chapter. In the field of justice, understanding the exact purpose of an algorithm is central to assessing the extent of rights violations it may cause. In civil law systems, judges must be able to understand what they are to evaluate, however complex. Expert reports clarify scientific evidence, but the judge remains the gatekeeper of the final assessment. This must remain true for AI systems. Technologists can explain how algorithms function, but it is for the judge to decide how much the technical element should weigh in the final judgment when personal integrity, freedom, equality, or human dignity are at stake.

One proposal worthy of consideration is the creation of specialized courts for disputes involving algorithms: "Perhaps in response to the complexity of algorithms, a court which focuses

on litigation involving algorithms will need to be developed. This is not a novel idea, as in the United States there is a Tax Court, a Court of Federal Claims, to name just a few specialized courts". Such a measure deserves evaluation if it allows for stronger protection of rights potentially endangered by AI.

It should also be noted that in 2018 the European Commission for the Efficiency of Justice explicitly called for transparency and user control in its European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment. The document sets out five guiding principles, of which the fourth and fifth require that data processing methods be accessible and understandable and subject to external audits (principle of transparency, impartiality, and fairness), and that users remain informed actors capable of controlling their choices (principle of user control). It further states that "a judicial system in keeping with its time would be one capable of establishing, administering, and guaranteeing genuine cyberethics for both the public and private sectors, insisting on total transparency and fairness in the functioning of algorithms, which may one day contribute to judicial decision-making".

Chapter 4

Case study: AI and big data in medicine

1. Introduction

This chapter focuses on the use of AI in medicine as a case study for several issues discussed earlier. The dialectic of opposites and the anthropomorphizing ideology that underlie AI also affect this specific field and merit closer examination. A critical analysis will follow of AI applications in both clinical practice and clinical research, exploring the multiple possibilities these technologies offer alongside their ethical challenges and corresponding legal responses.

The first point concerns the use of data and the conceptions underpinning it. Previous chapters have discussed the spread of an *essentialist* and deterministic ideology that shapes narratives about AI and the data it relies on. Borrowing Stefano Rodotà's words, one might say that a *mystical* adherence to data ideology has taken hold, rooted in an *infinite naïveté* about the significance of results obtained from data processing. Despite growing critical voices, the idea persists that data are the only source fully representing reality. This view becomes particularly problematic in medicine, where blind reliance on data risks diminishing the physician's role and undermining the current information model. The same "datadeology" influences the care relationship, creating dependence on intelligent technologies and limiting professionals' ability to use their own expertise to verify algorithmic analyses used for diagnosis, prognosis, and therapy. Maintaining a critical distance from technological adoption often justified as necessarily beneficial because it "saves lives" remains a challenge.

Medicine also reflects the anthropomorphizing narrative of AI, which will be exemplified in this chapter. The medical sphere is a privileged setting for observing how digitalization transforms human relationships. The care relationship has long been one of the most significant forms of human *relationality*, whose boundaries and roles have been discussed since Hippocratic times.

Originally, this relationship existed between doctor and patient. Only with the emergence of bioethics, driven by scientific and technological advances in the twentieth century, did it expand to include the medical team. Each actor has distinct roles, powers, and responsibilities from legal, ethical and deontological standpoints. Within this complex framework of intertwined ethical and juridical positions, AI now enters as a third agent. Its presence will gradually reshape the relational structure, either by improving medical work or by depersonalizing the care relationship. The dual-use nature of these technologies requires careful long-term assessment of their impact in medicine and of the purposes they should serve.

The digital transformation of medical practice carries risks of violating fundamental rights such as personal identity, non-discrimination, free development of personality, privacy, and dignity: risks that could outweigh the benefits of personalized medicine. Personalization, intended to make the right to health more effective, always has another side: the potential for increased control over individuals. Two models of personalized medicine can be outlined. The first keeps the physician central, aided by intelligent technologies to provide more individualized and accurate care. The second sidelines the physician, turning them into an executor of AI decisions. The latter model, often endorsed by holders of technological power inspired by transhumanist thought, poses an ethical challenge to the relational foundation of medicine. This relationship relies on mutual trust to sustain a virtuous collaboration between patients and health professionals, always oriented toward the patient's best interest.

The reflections that follow aim to highlight the critical aspects outlined here.

2. Data and their use in clinical practice and scientific research

Medicine is the field that produces the largest amount of highly sensitive personal data, including health, biometric, and genetic data, as well as those derived from neuroscientific research such as brain mapping. These data are essential to improving clinical practice and biomedical research. Today, their quantity is so vast that the term Big Data has been coined to describe this environment. Because of their scale and complexity, Big Data cannot be managed according to traditional data processing principles and require specific procedures to safeguard fundamental rights and personal information. Although definitions vary by discipline, there is general agreement that Big Data are characterized by the "4Vs": volume, velocity, variety, and veracity. These aspects and their ethical implications have already been discussed. Here, the focus is on their use in medicine.

Big Data include heterogeneous datasets originating directly from the medical field (genetic analyses, radiological images, medical records) as well as indirectly from other domains, such as lifestyle, environmental, socioeconomic, and behavioral data. The World Health Organization defines biomedical Big Data as part of a broader health data ecosystem. They are processed through algorithmic methods to train AI systems that can later provide diagnostic and prognostic support or generate databases for scientific research.

Their use raises ethical and legal issues that differ depending on whether data are employed in clinical practice or in research. In clinical settings, the potential offered by massive data availability is expressed in data-driven precision medicine, defined as the ability to predict and simulate individualized treatment strategies based on collected data within the framework of personalized, stratified, and precision medicine.

Despite these opportunities, significant ethical challenges persist. Many technical issues may gradually be solved through technological progress, but behind them lie deeper ethical assumptions about who holds decision-making authority. While AI technologies can enhance diagnostic accuracy, unresolved issues related to the 4Vs affect the reliability of diagnoses and the quality of care. Contrary to expectations of improving speed and precision, AI-assisted decisions can lead to diagnostic or therapeutic errors not attributable to medical negligence. Physicians, responsible for providing accurate information, must rely on unbiased outputs. This requires clean, high-quality data. Otherwise, even if AI is meant to assist and not replace the physician, it effectively does so, since decisions would depend on results beyond the doctor's control. This undermines the legitimacy of medical judgment and the duty to present accurate diagnostic and prognostic assessments.

Problems related to dataset composition are well known. Ensuring data completeness and quality is critical, especially when data come from or concern populations in low- and middle-income countries, where underrepresentation is endemic. Systemic bias remains in the data used to train AI systems, and additional difficulties arise from algorithmic models marked by opacity and limited explainability. These technical issues conceal ethical and legal risks, potentially violating principles and fundamental rights such as personal identity, non-discrimination, equality of treatment, privacy, and informational autonomy, including the right not to know. Altogether, they weaken a person-centered vision by rendering meaningful human control impossible. If human oversight is lost, assigning responsibility for medical errors becomes highly complex.

Another concern stems from the growing influence of dataism. Even the clinical setting risks reducing human beings to categories of risk or health status, neglecting individual diversity. By focusing on data, medicine shifts attention from the person to their numerical representation, risking a drift toward dehumanized or depersonalized medicine. This may lead to forms of data tyranny expressed in excessive control and monitoring beyond what health needs justify.

In scientific research, similar issues arise regarding data quality, validity, and completeness. Access to existing datasets is crucial for building new ones, but it raises technical and ethical problems, particularly concerning transfers between jurisdictions with differing data protection regimes. Beyond risks of theft or accidental disclosure, ensuring transparent and non-discriminatory data use is essential, especially for vulnerable groups such as minors, who could face future stigmatization based on data collected in childhood. These concerns have intensified since the Covid-19 pandemic accelerated the creation of health data infrastructures for diverse purposes. Datasets used to train AI still present reliability issues, as their freedom from bias cannot be guaranteed.

The use of large datasets in clinical practice and biomedical research thus opens unprecedented opportunities but also complex ethical questions that lack full legal definition. The proposed European Regulation on Artificial Intelligence is the first binding initiative addressing AI use in European society, though it does not explicitly cover the medical field. The future regulation will clarify its impact on medicine. If it remains genuinely human-centered, it could play an important role in safeguarding individuals and preventing the rise of a dehumanized medical model.

The risks of such dehumanization emerge most clearly in the care relationship, which will be examined in the next section.

3. AI and the care relationship: a third *super partes*?

In its Strategic Action Plan on Human Rights and Technologies in Biomedicine (2020-2025), published in November 2019, the Council of Europe's Committee on Bioethics (DH-BIO) included among its objectives the preparation of a report on the applications of AI in healthcare, with particular attention to the care relationship. In June 2022, its successor, the Steering Committee for Human Rights in the field of Biomedicine and Health (CDBIO), released The Impact of Artificial Intelligence on the Doctor-Patient Relationship. The report examined in depth the effects of AI on the care relationship, identifying its main critical aspects. Situating its analysis within the framework of the principles and provisions of the Convention on Human Rights and Biomedicine (Oviedo Convention), the Committee focused on AI as applied to biomedical research and patients. Although the report emphasizes that AI use in clinical contexts is still in its early stages, it offers an analysis of the potential transformation of the AI-mediated doctor-patient relationship. Building on some of the elements discussed in that report, this section adds further reflections, focusing mainly, though not exclusively, on cases in which AI systems play a role in medical decision-making, such as by providing diagnostic recommendations.

Before addressing these issues, it is useful to recall the key features of the care relationship as it has evolved since the second half of the twentieth century, largely thanks to bioethical reflection. This step is essential for understanding which aspects of this relationship AI may affect.

3.1. The care relationship between rights and duties

Today we take for granted the language of patients' rights, their decision-making autonomy regarding health, and the right to know (and not to know). But these are relatively recent achievements, the outcome of rights movements that emerged in the second half of the twentieth century in the Western world. These struggles involved groups historically denied decision-making autonomy – women, minors, and members of marginalized communities – who eventually demanded the ability to decide for themselves or, at least, to participate in decisions through proper access to information. Within this broader context, the patients' movement arose in the United States in the late 1960s and early 1970s and later spread across Western countries. A combination of cultural and economic conditions made these advances possible, the affirmation of personal autonomy as a principle, widespread economic well-being, and unprecedented scientific and technological progress. Together, they fueled the success of patients' claims. In an age where medicine gained the power to control vital and biological processes once left to chance, the right to make decisions, traditionally considered the exclusive competence of the physician (within a paternalistic model of care), could no longer be assumed.

Given the immense authority doctors hold, their role can no longer be defined solely in technical or scientific terms. In the new technological context, this view is insufficient to justify exclusive decision-making power, since it conceals ethical and moral issues that require recognition. It thus becomes necessary to move beyond the purely technical dimension of medicine to include its moral implications. While it remains undisputed that physicians must propose therapies and scenarios of care, and that their actions are bound by beneficence and non-maleficence, what has changed is the relationship between the physician's and the patient's ethical perspectives. In the paternalistic model, medical ethics prevailed unconditionally. From the 1970s onward, with the emergence of the liberal or informational model, the patient's ethics prevailed through the principle of autonomy. The final decision to accept or refuse treatment belongs to the patient, expressed through informed consent. In adhering to beneficence, the physician must

pursue not an abstract objective good but one defined through both the appropriateness of treatment and the evaluation of quality of life, which only the patient can truly determine.

The doctor-patient relationship thus becomes one founded on trust rather than asymmetry. The physician's central duty is to provide the information necessary for the patient to make an informed and autonomous decision. Bioethical scholarship has deeply developed this issue, clarifying the physician's informational role in relation to the patient's expression of informed consent. Beauchamp and Childress distinguished three stages in the process: disclosure (communicating information), understanding (ensuring comprehension), and informed consent (the patient's authorization).

Initially, disclosure was the main focus, especially in the United States, as a way to overcome physicians' reluctance to inform patients. However, it later became clear that revealing information is not enough unless the patient's understanding is also verified. Physicians must therefore assess whether patients truly comprehend what they have been told. Once understanding is ensured, the decision, whether to accept or refuse treatment, belongs fully to the patient, who must be given adequate time to reflect, especially when interventions are invasive or high-risk.

These ethical and moral principles have been translated into legal norms. In Europe, the first reference to the care relationship and to the tools that safeguard patient autonomy is the Oviedo Convention, which dedicates Chapter II to consent and advance treatment directives. The Charter of Fundamental Rights of the EU follows, stating in Article 3 that consent is the foundation of respect for personal integrity in the medical and biological fields. In Italy, Law No. 219/2017 provides an explicit legal framework for informed consent and advance directives.

According to established interpretation in jurisprudence, doctrine, and bioethics, informed consent legitimizes the physician's activity. The information provided must be complete (relative to the specific case), truthful, up-to-date, and comprehensible, tailored to the patient, not to an abstract standard. The patient may delegate information or fully entrust decisions to the doctor, but such a choice must be properly documented.

As emerges from these reflections, information and informed consent are the pillars of today's model of care. The doctrine of informed consent, developed since the mid-twentieth century, remains a central element of the doctor-patient relationship, and one that risks being profoundly altered by the introduction of AI as a decision-support tool. The next section will examine the main aspects to consider when the care relationship is mediated by AI.

3.2. The Care Relationship Mediated by AI

Several critical issues arise when analyzing how the care relationship may change when mediated by AI. For the purposes of this discussion, and in continuity with the previous analysis, attention will focus on two main aspects: the impact on the professionalism, competence, and responsibility of the physician, and the implications for the patient's decision-making capacity. The focus will remain on scenarios where AI is used as a support, not as a substitute, for human judgment, since this is the most immediate and realistic context.

Beginning with the physician's role, many challenges can be understood by examining the impact of AI on the two communicative phases central to medical ethics: disclosure (the transmission of information) and understanding (the patient's comprehension of that

information). Disclosure precedes any decision the patient must make. The information a doctor provides must include all elements necessary for autonomous decision-making, since without them, consent is neither morally nor legally valid. For example, if a given intervention carries a significant risk of permanent disability, the patient must be informed. But completeness does not mean exhaustiveness. Physicians are not required to transfer all their scientific knowledge to the patient but only what is necessary for an informed, uncoerced decision.

This raises key questions: How should the physician classify information concerning AI support? Must it always be disclosed? Are there situations where the physician may refrain from informing the patient? The answer depends on whether AI is used for diagnostic or prognostic purposes, or in surgical applications.

In diagnostic contexts, if AI is introduced as part of an innovative clinical process whose superiority over existing standards is unproven, its use lies at the boundary between clinical practice and research. It must therefore be justified by demonstrating that the innovation involves ethically acceptable risks. Physicians must be able to evaluate the AI system's validation criteria and understand both its potential and its limits. They should be capable of grasping the logic behind its predictions and determining whether this knowledge is relevant to diagnostic accuracy and, consequently, should be disclosed to the patient. This reinforces the need for AI systems trained on accurate and complete data, as well as for physicians to be involved in system design and familiar with the datasets that inform their decisions.

These steps are essential for managing both the disclosure of information and the patient's understanding of it, ensuring effective communication and genuine autonomy. Whenever knowledge of AI use is indispensable for an autonomous choice, the physician must also verify the patient's comprehension of AI's role in the decision-making process.

Many of these reflections also apply to the use of AI-assisted robotics in surgery. In this field, beyond technical preparation for explaining how the robotic system operates, the physician must also assess the patient's potential overconfidence or mistrust toward the technology and help address such perceptions. The anthropomorphizing ideology that machines are superior to humans because they are not affected by subjectivity runs deep in contemporary culture, and medicine is not exempt. This belief may lead both physicians and patients to overestimate machine reliability, underestimating the risk of error.

For physicians, this may result in overreliance, an excessive dependence on technology that can, over time, cause deskilling, or the erosion of clinical expertise. For patients, excessive trust may lead to unrealistic expectations and disillusionment if outcomes fail. Conversely, patients may distrust robotic systems and demand traditional procedures. In these cases, the physician must be prepared to explain the advantages and limitations of AI clearly. It remains unclear whether patients can insist that a surgical operation be performed exclusively by a human surgeon. If the robotic system represents an unvalidated clinical innovation, patients cannot be forced to undergo an experimental procedure. At the same time, the law states that physicians are not obliged to comply with requests that contravene legal, ethical, or professional standards.

During both the disclosure and understanding phases, it is important that the physician, possibly with legal counsel, clarify who would bear responsibility in the event of an error during surgery. This transparency is essential to preserve the trust relationship with the patient. The literature refers to this challenge as the problem of many hands, highlighting the ethical and legal difficulty of attributing responsibility when multiple actors are involved and roles are undefined.

Finally, one further concern must be noted: the risk of exacerbating, through *datification*, an abstract view of disease detached from the embodied person who suffers. This contradicts decades of bioethical reflection aimed at placing the patient, as a person with individual needs and beliefs, at the center of care. This risk is amplified by the transformation of the physician's role in the AI-mediated context, where doctors risk becoming specialized technicians treating the body as a machine, detached from the experience of suffering and care.

4. Neurotechnologies and Human Enhancement: The New Frontiers of Medicine

AI applications show great promise in medicine, particularly when combined with neurotechnologies. These are technologies that enable a direct connection with the brain, allowing neuronal activity to be monitored, recorded, or altered. Neurotechnologies that intervene on neural activity, whether open-loop or closed-loop, are generally divided into three categories: neurostimulation technologies, neuroprosthetic technologies, and brain-machine interfaces (BMIs or BCIs).

The first group includes devices using neural interfaces to stimulate parts of the central, peripheral, or autonomic nervous system. They may operate as open-loop or closed-loop systems. A notable example of the latter is neuromodulation therapy for Parkinson's disease, in which sensors adjust stimulation intensity to reduce tremors.

Neuroprosthetic technologies are devices designed to restore or replace cognitive, motor, or sensory functions. Cochlear implants, for instance, allow individuals with hearing loss to interpret sound signals through the stimulation of auditory neurons in the brainstem.

BMIs establish a direct link between the brain and an external device. These interfaces can be unidirectional, allowing users to control an external system or receive sensory input, or bidirectional, enabling both encoding and decoding functions. They are used, for example, in patients with paralysis to interpret brain signals and control robotic limbs, while restoring sensory feedback through targeted neural stimulation. BMIs are also employed in therapeutic contexts, such as deep brain stimulation (DBS) for conditions like depression, where they act directly on emotional states. Moreover, they are used to address cognitive deficits, such as those caused by Alzheimer's disease, by improving memory performance and even enabling the recovery of so-called flashback memories.

In sum, these technologies have already achieved, or promise to achieve, remarkable results in treating physical and mental disorders and improving patients' quality of life.

However, the same technologies also have potential applications beyond therapy. In 2009, the Panel for the Future of Science and Technology (STOA) published a major study on human enhancement, dedicating an entire section to deep brain stimulation technologies and their therapeutic and non-therapeutic prospects, along with related ethical concerns. This demonstrates that the topic has been under institutional consideration in Europe for over a decade. To understand the implications of neurotechnological developments linked to AI for non-therapeutic purposes, it is first necessary to outline the main coordinates of the complex debate on human enhancement.

4.1. Human enhancement in brief

The term human enhancement refers to interventions on the human body or brain made possible by scientific and technological advances, aimed at improving or increasing existing human abilities or creating new ones. Enhancing interventions can be therapeutic, when treating a pathology allows a patient to reach performance levels beyond normal standards, as in the case of prostheses enabling superior physical performance (the Pistorius case). However, the expression generally refers to non-therapeutic interventions on healthy individuals.

At the international level, the debate on human enhancement began in the 1960s in the United States, initially within sociology, as part of a broader reflection on the medicalization of society. In its early phase, the discussion focused mainly on the use of drugs – developed for therapeutic purposes – for enhancement instead. These include medications created to treat pathological conditions but used to increase cognitive abilities such as memory and concentration, or to improve physical performance, as in sports doping. Over time, the academic and institutional debate expanded to include the use of devices and drugs not only to enhance physical or cognitive capacities but also to intervene in individuals' moral traits. This is the controversial project of moral bioenhancement, aimed at improving the moral character of individuals to create a "better" society.

Human enhancement, in all its forms, raises multiple questions. Philosophical questions concern whether human nature should be seen as malleable and modifiable at will or as limited and fixed. Ethical questions involve the acceptability and justification of artificial improvement, the limits of acceptable risk, and whether what is "better" is necessarily "good" and therefore worth pursuing. Legal questions concern how the law should respond and how to balance competing interests fairly, whether through existing principles or by formulating new ones, as in the debate on neurorights.

From a philosophical standpoint, bioconservatives view human enhancement negatively, while transhumanists promote it, "the former determined to restore the rights of nature, the latter defending a new freedom: that of using without limits the unprecedented power we now possess". The topic lies at the crossroads of several crucial issues, particularly the preservation of the individual's intimate sphere and the principles protecting it at the normative level. Human enhancement touches upon several domains: 1) consumer rights, considering the growing market for direct-to-consumer devices used for recreational self-enhancement; 2) individual rights in military enhancement, often imposed rather than chosen; and 3) the rights of the person and the role of medicine.

Beyond both transhumanist enthusiasm and bioconservative alarmism, human enhancement is undeniably already part of medical practice. This is confirmed by Article 76 of the Italian Code of Medical Ethics (2014), explicitly dedicated to enhancement medicine. Yet, despite its inclusion in medical ethics, there is still no legal framework specifically regulating the matter. The increasing availability of enhancement tools for healthy individuals demands a legal response, particularly given the possibilities now offered by neurotechnologies.

Many neurotechnological therapies developed for medical use may eventually extend beyond that context. In some cases, medical intervention is required, such as implanting chips in the brain, but in others, it is not, as with wearable devices, some of which are already on the market. The following section offers brief ethical and legal reflections on both scenarios.

4.2. Ethical-legal profiles of neurotechnological enhancement

How should the prospects opened by neurotechnologies for non-therapeutic enhancement be evaluated? The assessment varies depending on whether neurotechnologies require medical intervention or are instead directly sold to consumers. It must also consider the degree of invasiveness of the technology, which ranges between two extremes: on one side, reversible or irreversible surgical interventions (for instance, brain chip implants or gene therapy), and on the other, wearable neurotechnological devices directly used by consumers.

Let us first consider invasive forms of neurotechnological enhancement, which necessarily require medical intervention.

These raise, first of all, questions about the acceptability of such enhancement within medicine. As seen earlier, opinions differ widely, but medical ethics has already addressed human enhancement, offering guidance to ensure a high level of patient protection.

Secondly, whenever a physician intervenes for invasive enhancement, the activity falls within the therapeutic relationship and is thus subject to all norms governing rights and duties in that context. This ensures the fundamental right to accurate information and control over that information, enabling genuine autonomy on the part of the individual requesting the intervention. Within this normative framework, the protection of this right finds its fullest expression. It is therefore desirable that medicine explicitly assume responsibility for these forms of enhancement, since the therapeutic relationship, based on mutual trust and the autonomy of both doctor and patient, can provide a safeguard against manipulation and abuse.

From a legal standpoint, forms of neurotechnological enhancement that connect the human brain with external devices have been analyzed through the lens of fundamental rights. While there is no explicit right to enhancement in the European context, there is a right to the free development of one's personality, which could, in theory, justify enhancement, including invasive forms. This extension, however, remains debated and lacks consensus within European ethical discourse. Beyond this, current discussions focus on protecting personal identity, integrity, and mental privacy against the possibility that neurotechnologies, particularly brain-machine interfaces, might allow external access to our thoughts, memories, and beliefs. Legal reflection here moves in two directions: some scholars argue that existing fundamental rights, interpreted broadly by courts, suffice to protect against such unprecedented intrusions, while others contend that new rights, neurorights, are urgently needed. The demand for neurorights has already reached the constitutional level in some countries. Chile, for example, has proposed amending its Constitution to include the protection of mental integrity under Article 19, paragraph 1.

If we move from neurotechnologies requiring medical intervention to those directly marketed to consumers, devices used without professional mediation, we see that, in this context, the absence of a trust-based relationship with an expert leaves many ethical and legal issues unresolved. These include: 1) How can consumers of neurotechnologies for entertainment or relaxation be provided with adequate information about potential risks of unsupervised use? 2) How can it be ensured that such information is not only received but also understood, enabling genuinely autonomous decisions? 3) Given the growing market for brain data, how can consumers be protected from economic forces that influence both markets and institutional decisions regarding the commercialization of these devices?

Possible answers to these questions may follow the recommendations below, though a deeper legal analysis will be necessary to equip policymakers with suitable regulatory tools.

General recommendations, not exhaustive but useful for shaping the legal framework, include:
- establishing institutional communication channels to provide accurate information about the risks and benefits of enhancement-oriented neurotechnologies, allowing the public to interact with experts, with adequate advertising of these channels;
- ensuring the possibility of consulting a qualified physician in case of side effects from wearable neurotechnologies;
- guaranteeing transparency about the commercial interests driving large-scale diffusion of these technologies;
- designing and building neurotechnologies ethically, especially those capable of accessing or influencing data and thoughts stored in the brain, as they expose individuals to risks of personality manipulation.

4.3. Neurotechnologies and the anthropomorphizing ideology: some concluding reflections

The recommendations above are essential to counter both the logic of opposites and the anthropomorphizing ideology, which finds in the convergence of AI and neurotechnologies a powerful ally.

Current brain-machine interfaces can, in some respects, be seen as heirs to the cyborg concept, developed in the last century within medical research. The term originated in a biomedical context, specifically in NASA's space research between the late 1950s and early 1960s, coined by two physicians, Clynes and Kline, who experimented with mechanical devices implanted in astronauts' bodies to inject biochemically active substances that could overcome physical and psychological limits caused by space missions in zero gravity.

Today, transhumanists consider the cyborg a transitional stage between the human condition, limited and imperfect, and the full post-human condition, in which the body will no longer be necessary. The phenomenon of the Metaverseexemplifies this evolving conception and the massive commercial drive behind transhumanist philosophy. It is believed that one day it will be possible to upload our brains into such a virtual reality, achieving a form of eternal life. At the core of this view lies a paradoxical assumption: that humanity, reshaped by the very machines it has created, will become an improved version of itself. From this follows the idea that humans must use technology to pursue this improvement to its ultimate consequence, renouncing corporeality and making their minds open and accessible to all.

This transformation of humans into machines will, for some time, require the cooperation of medicine, which explains the transhumanist push for medicine to embrace enhancement without resistance.

The transhumanist proposal is deeply prescriptive, as it seeks to define what is "good" in absolute terms and therefore must be pursued. It identifies which traits should be abandoned and which developed, assuming universal agreement on these choices. By exploiting humanity's primal fears – death and aging – transhumanism presents itself as a salvific alternative offered by technology. Implicitly, it prescribes a perfect life designed by a few, without consultation. As observed, "with transhumanism, a new religious paradigm emerges: that of Man-God. It is no longer the renunciation of the atheist who sees himself alone in the universe, but the proud affirmation of all that man can do, including creating life and recreating himself".

Such a prescriptive stance cannot be reconciled with the axiological framework of fundamental rights, conceived as limits to concentrated power and as guarantees for the most vulnerable. It stands in tension with the transhumanist idea of unregulated innovation, or rather, innovation governed only by market logic, which disregards the need to cultivate autonomy and self-determination through institutional, cultural, and legal means.

If we are to be protagonists of innovation and not its passive subjects, we must recognize that ethics and law are the essential instruments for counterbalancing the growing power of the few over the many. They are indispensable to reaffirm that fundamental rights define the *sphere of undecidable*, beyond market logic, and to restore the centrality of human dignity enshrined in constitutional systems.