T

Silvia Dadà

From Risk Society to Digital Risk Society: Systemic Risk and the Challenges of the EU AI Act

1. *History of the concept of risk*

The notion of risk has undergone numerous transformations over the centuries, while preserving its fundamental function: an attempt to control and manage the disorder inherent in reality. This dimension was particularly emphasized by Mary Douglas, an anthropologist who explored the political and cultural significance of the concept of risk. Closely tied to the themes of contamination and purity, risk belongs to that set of symbols through which societies assign responsibility, identify blame, and create social cohesion:

> Whose fault? is the first question. Then, what action? Which means, what damages? what compensation? what restitution? and the preventive action is to improve the coding of risk in the domain which has turned out to be inadequately covered. Under the banner of risk reduction, a new blaming system has replaced the former combination of moralistic condemning the victim and opportunistic condemning the victim's incompetence (Douglas, 1992: 15-16).

Humanity has always confronted the future and its uncertainty, though not all cultures have adopted the same strategies for doing so. In antiquity, for instance, concerns over unpredictability were entrusted to the relationship with the divine and to the interpretation of divine will, without recourse to the notion of risk or to the techniques of calculation, management, and assessment associated with it. As Peter Bernstein (1996) observes, the turning point lies in the shift from reliance on the gods to reliance on rational and strategic calculation, which becomes the defining feature of the modern notion of risk. There is, therefore, a close connection between the concept of risk and the process of secularization. As Ulrich Beck puts it: «When

 DOI: 10.4454/3cz9p720

Nietzsche announces: God is dead, then that has the – ironic – consequence that from now on human beings must find (or invent) their own explanations and justifications for the disasters which threaten them» (Beck, 2008).

Although the origins of the term remain uncertain (possibly Arabic), the Neo-Latin form *risicum* is already attested in the medieval period (Luhman, 1996). Its earliest uses are primarily found in connection with maritime trade and ship insurance, where it referred to adverse events – such as floods, epidemics, or earthquakes – that could jeopardize the success of a voyage. These phenomena were not dependent on human action, nor were they considered calculable. This excluded any connection between this early idea of risk and human responsibility. The scope for action and prediction was thus extremely limited. In this initial phase, risk essentially denoted the danger of suffering misfortune, over which human intervention could exert little influence.

With the advent of modernity in the eighteenth century, however, the meaning of risk expanded to encompass elements attributable to human agency. Within this context, risk acquired a calculable and objective character. The distinction between risk and danger became more sharply defined: «[…] in the case of risk/danger in the fact that only in the case of risk does decision making (that is to say contingency) play a role. One is exposed to dangers. Of course, the behaviour of those concerned» (Luhmann, 1993: 23).

A series of factors contributed to the development of this new understanding. The Enlightenment fostered a general faith in human progress and in the rational, objective comprehension of the world. Moreover, advances in the fields of probability and statistics made it possible to refine techniques of prediction on a mathematical basis[1].

The modern sense of "risk" is therefore characterized by its distinction, on the one hand, from "danger" (since it depends on human decision), and on the other, from "uncertainty" (because of its rational and measurable nature). It is also defined by its ambivalence, encompassing both positive and negative connotations. Especially in financial and insurance contexts, the assumption of risks may indeed entail loss, but it may equally represent opportunity and gain.

[1] On the relationship between risk and statistics see Bernstein (1996).

2. *Global risk in the risk society*

Things change significantly in the transition from modernity to postmodernity, or late modernity, the stage that marks the entry into what Ulrich Beck aptly called the "risk society" (1992). With the growing mistrust of scientific certainty and objectivity, the fragmentation of grand narratives and traditions (accompanied by suspicion toward authority and institutions), the concept of risk gradually shifts from the dimension of control to that of uncertainty. As Anthony Giddens argues:

> If risk has always been conceived as a way of dealing with the future, of managing it and bringing it under our control, this is no longer the case today: our attempts to control the future tend to rebound against us, forcing us to consider alternative ways of engaging with uncertainty (Giddens, 2000: 40, our translation).

A pervasive sense of insecurity troubles society, primarily because the very status of the dangers we face is changing. Deborah Lupton, (1999) in her work dedicated to this topic, identifies six types of risks that characterize our time: (1) environmental risks; (2) lifestyle risks; (3) health risks; (4) risks in interpersonal relationships; (5) economic risks; and (6) risks of crime. Although these risks manifest in distinct domains, their common point of origin lies in their connection with the development of new technologies. As Luhmann (2002) emphasizes, technology "transforms dangers into risks simply because it creates possibilities for decision-making that previously did not exist". Thus, one can distinguish between *technical risks* – adverse events caused by the structure of the technology itself (for example, an accident due to a system malfunction) – and *sociotechnical risks*, which result from the deliberate use of technology and the power relations guiding such use (for example, the detonation of a bomb).

It is therefore no coincidence that this evolution of the concept of risk has become even more evident in our era of rapid technological progress. On the one hand, calculating and predictive capacities – as well as life-enhancing tools – have expanded considerably; on the other, margins of instability and unpredictability regarding the consequences of human action have increased proportionately. Even the distinction between natural and artificial is blurred: the looming catastrophes threatening our planet are increasingly caused by human activity and the use of technology, to the point that François Ewald has spoken of genuine "artificial catastrophes".

In Beck's postmodern society, risks acquire a *global* character (Beck, 2008) which comprises three main features: (1) the *delocalization* (spatial,

temporal, and social) of their causes and consequences; (2) the *incalculability* of their outcomes and impacts, rendering every decision grounded in a fundamental not-knowing; and (3) their *non-compensability*, which makes the logic of indemnification impossible.

These characteristics make calculability and control especially difficult, since human action now exerts effects so extensive across space and time that the causal chain of responsibility is often uncertain and difficult – if not impossible – to reconstruct. Responsibility for present action increasingly conceals the threat of unimaginable and unforeseen consequences. In this context, risk becomes uncertainty, undermining the possibilities of compensation, damage limitation, security, and calculation: «Risk society is a catastrophic society. In it the exceptional condition threatens to become the norm» (Beck, 1999: 24). In our age, therefore, the weight and importance of each decision is magnified, as is the desire to identify those responsible. Yet this identification grows ever more difficult, due to both uncertainty and systemic complexity. The specialization and division of labor in industrialized society have fragmented responsibility to the point of near dissolution. As Anders perceptively observed:

> The 'technification' of our being: the fact that today it is possible that unknowingly and indirectly, like screws in a machine, we can be used in actions, the effects of which are beyond the horizon of our eyes and imagination, and of which, could we imagine them, we could not approve – this fact has changed the very foundations of our moral existence. Thus, we can become 'guiltlessly guilty', a condition which had not existed in the technically less advanced times of our fathers (Anders, 1962: 1).

Thus, the modern idea of risk gives way to a far more unstable and elusive concept, one that no longer allows us fully to contain and control our future and the consequences of our actions, though it still strives to «calculate the incalculable» (Dean, 1999). As Mark Coeckelbergh (2015a) has argued, risk – or more precisely, *being-at-risk* – becomes an existential category, describing humanity's condition of exposure and uncertainty in the age of new technologies.

The purpose of risk forecasting and calculation has always been to devise strategies for eliminating or mitigating threats. Covello and Mumpower (1985) suggest that, historically, the main techniques have been: (1) avoiding risk through prohibitions; (2) regulating and modifying human activities to reduce the magnitude of risk; (3) reducing the vulnerability of exposed populations; (4) developing interventions after events to mitigate impact; and (5) compensating for damages through insurance mechanisms.

Today, however, in the face of such unprecedented power and unpredictability, our risk management techniques – as Beck vividly puts it – resemble bicycle brakes mounted on an intercontinental rocket. Precautionary measures for prevention and damage reduction often prove ineffective, since the catastrophic consequences of our actions are not fully foreseeable (consider pollution and climate change), while *ex ante* solutions such as intervention and insurance appear neither sufficient nor proportionate. After all, how and whom can we compensate in the face of an irreversible global catastrophe, such as the explosion of a nuclear power plant?

Beyond its historical evolution (with the significant shifts in meaning already traced), the concept of risk also embodies a plurality of interpretations. Although the landscape is broad and varied, two principal approaches may be distinguished: one that conceives risk in an essentially objective and rational way, and another that views it as a social and cultural construct, not dependent on factuality itself but on how it is managed and represented.

The first perspective is adopted primarily by disciplines such as engineering, actuarial mathematics, statistics, and epidemiology, and is characterized by a technical-scientific approach. Here, risk is understood as something inherent in reality itself; what is subjective is not the risk but rather its perception, which differs between experts and laypeople and is influenced by cognitive mechanisms that undermine rational evaluation. Risk is thus treated as an objectively measurable fact, the product of the probability of an event and the severity of its consequences: magnitude ($R = P \times D$). Attempts at containment and mitigation focus on altering one of these two variables. Both the identification of risk and the search for solutions are generally reduced to technical factors.

The second perspective, by contrast, is sociocultural in nature and developed largely by philosophical, anthropological, and sociological reflection. Its basic assumption is that risk, before being an objective datum, is the interpretive category through which modern society is governed, and by which it defines its relationship with otherness, with knowledge, and with the future. This confers upon the category of risk (and upon the decisions deriving from it) a political dimension, whereby specific strategies of governance and responsibility distribution are elaborated in response to risks. While it is important to distinguish between the various interpretations, as well as between risk and its perception, it must nevertheless be recognized that the two remain deeply interdependent: «scientific rationality without social rationality remains empty, but social rationality without scientific rationality remais blind» (Beck, 1992: 30).

3. *Systemic Risk*

The transition to the late-modern era has thus given rise to a different conception of risk, which Beck has defined as *global*, insofar as it is de-localised, incalculable, and non-compensable. Within this theoretical framework, the early 2000s saw growing scholarly attention to a more specific dimension of risk: *systemic risk*. This concept, widely adopted across various scientific and disciplinary domains, offers a more nuanced understanding of how certain contemporary risks have emerged and evolved. To grasp this concept, it is first necessary to briefly clarify the meaning of "system". Drawing on philosophical reflection, cybernetics, and complex system science, Terenzio (2025) offers a detailed analysis of the term, identifying its core features. Chief among these is the relational nature of its elements, articulated through a dynamic interplay between parts and the whole. A system thus emerges as an organized and autonomous configuration of components, sustained by nonlinear dynamics. Yet, the order generated within systems is neither absolute nor permanent; rather, it constitutes a fragile equilibrium, continually exposed to disruptions and emergent phenomena capable of transforming individual elements and reconfiguring the system as a whole. It is precisely this interconnection and inherent instability that give rise to what are known as systemic risks.

It is challenging to establish a single, universally accepted definition of this concept, given the breadth of its theoretical and practical applications. One of the most widely cited definitions is offered by Kaufman and Scott (2003, p. 372) who state: «Systemic risk refers to the risk or probability of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, and is evidenced by co-movements (correlation) among most or all parts». A defining feature of systemic risk, therefore, is its tendency to affect multiple components of a system simultaneously. In a similar vein, the definition provided by the Organization for Economic Co-operation and Development introduces the concept of systemic risk as referring to those "risks that threaten society's essential systems, such as infrastructure, health care and telecommunications" (OECD, 2003). A key emphasis in these studies lies in the transmission and scope of risk across the interconnected components of the system (Poledna *et al.*, 2020)[2].

[2] This does not imply that the scope of risk is necessarily global. As Aven and Renn (2019) emphasize, the extent of risk may be regional, national, or global. What is crucial, however, is the internal relationship among the components of the system, regardless of its geographical scale.

Systemic risks can be understood as threats whereby localized failures, accidents, or disruptions have the potential to affect an entire system through contagion mechanisms. In highly interconnected environments – such as health, ecosystems, infrastructure, and food – characterized by complex causal structures and non-linear relationships between causes and effects risk does not remain confined to isolated events but spreads across networks of mutual dependence. The limited understanding of these interconnections, combined with the inherent complexity of such systems, poses significant challenges to both prevention and effective risk management. Systemic risk is characterized by cascading effects that spread across interconnected systems through the movement of people, goods, capital, and information across borders. These dynamics can lead to widespread disruption or even systemic collapse.

The paradigmatic example of systemic risk – indeed, the event after which the very term gained widespread currency – is the global financial crisis of 2007[3]. Prior to this crisis, banking regulation was largely microprudential in orientation, concentrating on individual institutions and the risks apparent in their balance sheets. This approach rested on the assumption that constraining excessive risk-taking at the level of each bank would suffice to prevent the build-up of systemic vulnerabilities across the financial system. The crisis, however, exposed the shortcomings of this framework, demonstrating its inability to capture the intricate interconnections among financial institutions and markets (Allen and Carletti, 2013). More recently, the COVID-19 pandemic (Trump *et al.*, 2021) has triggered a global crisis, clearly revealing the deeply interconnected nature of key systems such as healthcare, finance, food supply chains, and labor markets. This event underscored how disruptions in one domain can rapidly cascade across others, amplifying vulnerabilities and challenging the resilience of complex socioeconomic structures.

Renn *et al.* (2022, 1904-1905) identify several defining properties of systemic risks. First, *complexity*, which refers to the difficulty of tracing and reconstructing the causal relationships within the system. Second, *uncertainty*, stemming from the indeterminacy of causes and characteristics of the phenomena, which in turn undermines confidence in both analysis and decision-making. Third, *ambiguity*, understood as the coexistence of mul-

[3] Other illustrative examples of these dynamics include the desertification and collapse of the Aral Sea, pandemics, cybersecurity threats, global climate change, and the depletion of fish stocks (IRGC, 2018).

tiple, and sometimes conflicting, interpretations of the same phenomenon. Finally, the *ripple effect,* whereby a disruption generates cascading consequences that extend far beyond the initial source of risk.

A valuable overview of systemic risk research is provided by a recent Briefing Note of the UN Office for Disaster Risk Reduction (Sillmann *et al.*, 2022), which offers an integrated perspective on climate, environmental, and disaster risk science. It examines the evolution of systemic risk across disciplines, identifying common conceptual threads without enforcing a singular definition. The Note outlines key attributes of systemic risk and discusses the types of data and information needed to enhance its practical understanding. In the report, systemic risks are examined along several dimensions that help capture their complexity and scope. The first concerns their *scale*: such risks may unfold at the global, national, regional, or local level, yet in all cases they tend to generate repercussions that transcend their initial boundaries. A second dimension relates to the *relations* within and across systems, characterized by feedback loops, interactions, and interdependencies. These intertwined connections amplify the potential for risk propagation and make it difficult to address threats in isolation. Third, systemic risks are marked by the *comprehensibility of the system*. They often involve a lack of knowledge, unpredictability, uncertainty, and ambiguity, which may lead to an underestimation of both their causes and their potential consequences. This opacity undermines the ability to anticipate developments or to design timely responses. A fourth dimension concerns the *transboundary effects* that such risks produce. These include contagion, cascading dynamics, and non-linear processes, through which a localized disruption can trigger a ripple effect, spreading across sectors and jurisdictions that initially appear unrelated.

Finally, systemic risks are defined by their *possible outcomes*, ranging from breakdowns and disruptions to the collapse of entire economic, social, or environmental systems. It is this capacity to destabilize or disintegrate complex structures that makes systemic risks particularly difficult to prevent and to govern.

When considered in light of these dimensions, the concept of systemic risk emerges as a powerful analytical category that extends and refines the sociological theorization of the "risk society" proposed by Beck. Systemic risk does not simply supplant the conventional risks of industrial modernity with those of a globalized and technologically mediated world; rather, it reconfigures the landscape of risk by juxtaposing two coexisting dimensions. The first comprises risks that remain bounded, relatively predictable, and

therefore more susceptible to regulatory control and mitigation strategies, such as regulation and control mechanisms[4]. The second consists of risks that are profoundly complex, marked by interdependencies, cascading effects, and epistemic opacity, which resist both prediction and governance.

From this perspective, systemic risk can be seen as a conceptual bridge between Beck's diagnosis of the reflexive modernity of late industrial societies and contemporary debates on global vulnerabilities, interconnectivity, and the fragility of socio-technical systems. It highlights how the dynamics of globalization, digitalization, and environmental change have intensified the conditions that Beck identified, producing a risk environment that is not only more pervasive but also qualitatively different. Thus, systemic risk serves not merely as a descriptive category but as a normative and operational framework, compelling institutions to rethink strategies of governance, resilience, and adaptation in a world where uncertainty is not an exception but a constitutive feature of social life.

4. *Digital risk society*

After reconstructing the notion of risk from antiquity to the contemporary age, we can now examine how the concept applies to the risks associated with AI systems, focusing on their specific characteristics and the ways in which they are addressed.

It is widely acknowledged that artificial intelligence (AI) poses significant risks. These threats can be categorized according to various criteria. The study by Hendrycks *et al.* (2023) provides an overview of the main sources of catastrophic risk associated with artificial intelligence, organizing them into four categories: 1) *malicious use*, the deliberate deployment of AI systems to cause harm; 2) *AI race*, the adoption of unsafe systems or the relinquishment of human control to machines driven by competitive pressures; 3) *organizational risks*, the increased likelihood of catastrophic failures resulting from system complexity; 4) *loss of control*, the inherent difficulty in governing agents that may become significantly more intelligent than humans.

These risks stem both from the *inherent nature* of AI models – particularly those developed through machine learning algorithms – and from their

[4] Common examples include bicycle theft, foodborne illnesses, and traffic accidents – events that, while harmful, remain localized and can be effectively addressed using existing tools and procedures.

potential misuse. For example, AI can be employed for real-time individual recognition or user profiling, raising serious ethical and privacy concerns.

The first category concerns *technical risks*, namely errors in outputs linked to the quality of data and the reliability of algorithms. These issues are primarily related to accuracy and robustness. Since such systems are fueled by vast quantities of data – often gathered from our online activities – it is essential that datasets faithfully reflect the reality they are meant to represent. If datasets lack accuracy, consistency, completeness, or correctness, the resulting outputs will be flawed, and when used for decision-making, they may expose individuals to significant harm. Likewise, the algorithms processing these datasets must operate according to logical procedures aligned with their intended purposes. Yet how can the accuracy of data and algorithms be verified? The greatest difficulty lies in detecting and identifying errors, especially in systems based on deep learning and neural networks, which are often opaque and difficult to interpret. This opacity problem exacerbates the challenge, as it prevents human understanding of outputs and their validity.

A second category comprises *sociotechnical risks*, which arise from the interaction between humans and AI systems. One notable example is cyberattacks, where generative AI is exploited for phishing, malware distribution, deepfakes, and other forms of social engineering. Sociotechnical risks also emerge from the legitimate use of AI, where certain applications may nonetheless expose individuals or groups to vulnerabilities – for instance, biometric tracking and facial recognition technologies, which may compromise individual freedoms as a side effect of their deployment[5]. Both technical and sociotechnical risks ultimately affect human beings, but their causes differ: the former stem from technical deficiencies, while the latter arise from the human-machine relationship.

Beyond this distinction based on causation, risks may also be classified by their object. One major category concerns privacy risks, related to the management of personal data, which may be used by companies for purposes beyond those to which individuals have consented. Such practices blur the traditional boundary between public and private life by enabling the disclosure of information – such as health status, financial standing, or political preferences – that was once strictly personal. This proliferation of data can create a pervasive sense of surveillance, as individuals are con-

[5] In their study, Guan and colleagues (2022) distinguish between «technical risk» and «management risk», defining the former as encompassing «algorithm risk, data risk, and technology risk», and the latter as including «management risk and decision risk».

tinuously monitored through their devices (computers, smartphones, smartwatches) (Lupton, 2016).

Human autonomy is also at stake, since increasingly sophisticated profiling techniques influence our choices, not only in consumer behavior but also in political opinion-formation. The Cambridge Analytica scandal demonstrated the severe consequences of such practices, but AI is now routinely employed for electoral and propagandistic purposes (O'Neill, 2018; Hinds *et al.*, 2020). This trend is especially visible in online communities – our new digital *agorà* – where opinion exchange has become polarized and self-referential, generating phenomena such as filter bubbles (Parisier, 2011) and echo chambers (Cinelli *et al.*, 2020). These dynamics threaten democratic deliberation processes and compromise the quality of information, which is increasingly polluted by fake news and other forms of truth manipulation (Giusti and Piras, 2020).

Another pressing issue is algorithmic bias, which undermines fairness and justice in automated decision-making. Biases may originate from humans who embed them into systems, or they may be produced autonomously by the systems themselves. Datasets often conceal gender, ethnic, economic, or cultural prejudices that can profoundly harm vulnerable groups. A notable example is the use of the COMPAS software in the United States criminal justice system, which, in assessing recidivism risk, systematically assigned higher scores to African American defendants (Angwin *et al.*, 2023). Similar issues have arisen in medicine, where diagnostic accuracy for minority populations has proven lower due to underrepresentation in training data, leading to discriminatory outcomes (Guerrero *et al.*, 2018).

It is also necessary to consider the so-called *existential risks* (Cappelen *et al.*, 2025) – future scenarios in which AI becomes extremely powerful, evolving into a form of "superintelligence" (Bostrom, 2014) capable of subjugating or even destroying humanity. While some view this possibility as an unlikely dystopian projection, the rapid advancement of generative AI compels us not to dismiss such concerns lightly.

Finally, AI generates *externalities* (Hagendorff, 2022), i.e., consequences seemingly unrelated to its use but directly stemming from it, particularly affecting the environment and labor. Data development, collection, and storage entail a significant ecological footprint, exacerbating the planet's precarious condition. In the labor market, AI displaces human work in specific sectors, with some professions expected to disappear, while others – such as those of so-called *crowdkworkers* – emerge in precarious conditions with little legal protection.

This brief overview of the new threats associated with AI invites reflection in relation to our broader theme. We observe both elements of continuity with past risks and significant elements of novelty. As in earlier times, risks often assume a global dimension and may lead to catastrophic outcomes. In the case of AI, such globalized risks derive primarily from the massive scale of data involved and the pervasiveness of the technology.

Sundberg (2023) conceptualizes this emerging landscape as the *digital risk society*, a context defined by the pervasive presence of intangible technologies which, although frequently presented as solutions, simultaneously generate new and multifaceted risks. This society is further characterized by processes of dehumanization, as an ever-growing range of tasks once carried out by humans is increasingly delegated to machines whose internal operations remain opaque.

Since AI systems are embedded in nearly every aspect of daily life, we face what Mark Coeckelbergh (2015b) calls the «tragedy of the master». If technology was originally conceived – following Aristotelian logic – as a servant of human purposes, today our dependence has deepened to the point of entangling us in a Hegelian master-slave dialectic, defined by dependence and alienation:

> There is a risk that the automation technology we developed and use to serve us renders us vulnerable and dependent in new ways, creates distance between us and material reality, and "automates" us in the sense that we have to adapt our practices to what automation technology does and can do (Coeckelbergh, 2015b: 222).

As with the risks of the twentieth century, those arising from AI can be described as incalculable and unpredictable, thus belonging more to the realm of uncertainty. Unlike the past, however, this unpredictability does not stem solely from human inability to foresee consequences across temporal and spatial distances. In AI systems, often characterized by opacity, it is the functioning of the system itself that remains unknown, presenting a "black box" to human users. This becomes especially problematic when algorithmic decisions are applied in sensitive contexts such as finance or healthcare, where we must rely on outputs that are unintelligible yet generate new forms of correlation. AI risks are therefore exponentially incalculable and unpredictable.

Another novel aspect is that the actions of AI systems – and especially their outcomes – do not necessarily depend on human decisions (Fabris, 2018). While industrial automation has existed in the past, today we increasingly rely on algorithmic decision-making. The ability of AI systems to

learn and independently generate decisions introduces an additional layer of autonomy. This raises profound ethical and legal questions regarding responsibility, but it also reshapes the very configuration of risks. If the distinction between risk and danger rests on whether adverse events derive from human decisions or occur independently of them, then AI-related harms should more accurately be classified as dangers. This shift challenges traditional notions of agency and accountability.

In conclusion, AI systems are *experimental technologies* (van de Poel, 2016) which are those technologies whose risks and benefits are hard to estimate before they are properly inserted in their context of use. The concept of risk in relation to AI is closely tied to ideas of catastrophe, uncertainty, and danger, owing to its global reach, its incalculable and unpredictable character, and its independence from human decision-making. Addressing these challenges requires a dual approach: technological innovation and regulatory oversight. A global research community is actively engaged in developing technical solutions to mitigate algorithmic risks. At the same time, comprehensive legal frameworks are essential to ensure responsible AI deployment.

5. *The Risk-Based Approach in the AI ActFine module*

In response to the rapid expansion of the *digital risk society*, recent years have witnessed growing attention to the ethical and regulatory dimensions of artificial intelligence. Following a series of *soft law* initiatives aimed at guiding its responsible development (Fabris *et al.*, 2024), the need for a more robust and binding regulatory framework became evident – one capable of clearly delineating the boundaries of AI system design, production, and use. In the case of the European Union, this process culminated, after three years of debate and drafting, in the adoption of the *AI Act* in the summer of 2024 (Casonato and Olivato, 2024). This document represents a groundbreaking step at the global level, serving as a source of inspiration for other legislative initiatives and exerting a significant influence on the broader market well beyond the boundaries of the European Union (Bradford, 2020).

In this document, the concept of "risk" plays a central role, appearing over 300 times throughout the text.

Article 3 of the AI Act provides a set of definitions that serve as a useful framework for interpreting the remainder of the text. The Act introduces the notion of "risk", underscoring how central this concept is to the overall structure of the regulation:

2. 'risk' means the combination of the probability of an occurrence of harm and the severity of that harm.

As we can see, this definition reflects a modern understanding of *objective risk*, conceived as the product of the likelihood of an event occurring and the magnitude of its potential damage.

As stated in Recital 26, the principal strategy underpinning the AI Act – commonly referred to as the *risk-based approach* – constitutes a comprehensive response to the diverse risks arising from the deployment of AI systems within the European single market. Furthermore, it establishes a proportionate and balanced regulatory framework designed to address the principal challenges that characterize the contemporary technological environment:

> In order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk-based approach should be followed. That approach should tailor the type and content of such rules to the intensity and scope of the risks that AI systems can generate. It is therefore necessary to prohibit certain unacceptable AI practices, to lay down requirements for high-risk AI systems and obligations for the relevant operators, and to lay down transparency obligations for certain AI systems (Recital 26).

To ensure a proportionate regulatory approach that safeguards both fundamental rights and the integrity of the market while fostering technological progress, the AI Act introduces a distinction among different levels of risk, conceptualized as a "risk pyramid". Four categories are identified *ex ante*: (1) unacceptable-risk systems, (2) high-risk systems, (3) limited-risk systems – primarily concerning transparency obligations – and (4) low- or minimal-risk systems.

Under Article 5 of the AI Act, the "unacceptable risk" category encompasses AI systems deemed fundamentally incompatible with EU values and rights, and are therefore prohibited. These include applications that manipulate individuals through subliminal techniques or exploit vulnerabilities based on age, disability, or socioeconomic status, thereby impairing autonomy and causing potential harm. The Act also bans AI systems used for social scoring, predictive policing based on profiling, and indiscriminate collection of facial images from online sources or surveillance. Furthermore, emotional recognition technologies in workplaces or educational settings are prohibited unless justified by medical or safety needs. Biometric categorization based on sensitive attributes – such as race, political affiliation, or religious beliefs – is likewise forbidden. Real-time remote biometric identification in public spaces is only permitted under narrowly

defined law enforcement conditions, subject to judicial oversight.

High-risk AI systems – whose identification and regulation constitute the core focus of the document, beginning with Article 6 – are those that, in order to be placed on the market, must comply with a set of specific obligations. According to the AI Act, high-risk artificial intelligence applications are those that may negatively impact human safety and security, fundamental rights, or the environment. In order to be placed on the European market, these systems must comply with specific requirements (Articles 8–27), including the implementation of risk and quality management systems, appropriate data governance mechanisms, and the use of relevant, representative, and error-free datasets. They must also provide technical documentation, ensure the automatic logging of significant events, and enable effective human oversight throughout their operation.

The scope of high-risk systems covers products already regulated under EU law, such as medical devices, toys, radio equipment, vehicles, lifts, and civil safety systems. Furthermore, Annex III extends this classification to domains including critical infrastructure, education and training, employment and labor management, access to essential private and public services, law enforcement, migration, asylum and border control, as well as the administration of justice and democratic processes.

The third level concerns risks arising particularly from chatbots, deepfakes, and other forms of AI-generated content, where it may be difficult to discern whether one is interacting with a human or a machine. To mitigate such risks, the AI Act imposes specific transparency obligations. The final level represents a residual category, encompassing all systems not included in the previous tiers. These systems are not subject to binding obligations but may voluntarily adhere to soft law instruments, such as codes of ethical conduct and guidelines.

The reference to the *acceptability of risk*, determined through a cost–benefit analysis, aims to assess the degree of trustworthiness of AI systems (Fraser, Bello y Villarino, 2023). The allocation of these risk levels depends primarily on the context of application. However, the rapid development of technologies has highlighted the need for oversight that goes beyond merely identifying applications. Attention must now also be directed toward the underlying AI models themselves, prior to their deployment within specific systems. This approach is particularly crucial given the widespread adoption of generative AI and large language models, which complicate the task of defining discrete applications (Novelli *et al.*, 2023). For this reason, the original risk pyramid – introduced in the first regulatory proposal of April

2021 – has been supplemented by a new dimension, encompassing general-purpose AI models and the potential systemic risks they may generate.

As stated in Article 3, definition 63, a "general-purpose AI model" refers to models trained on large volumes of data, characterized by a high degree of generality and capable of competently performing a wide range of distinct tasks. These models can be integrated into various systems or applications and are able to execute functions such as image and speech recognition, audio and video generation, pattern detection, question answering, and text translation. The type of risk associated with such models is defined as "systemic", as outlined in definition 65 of the same article:

> 'systemic risk' means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a *significant impact* on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society *as a whole*, that can be propagated at scale across the value chain.

This second definition, although admittedly broad and somewhat indeterminate in scope, elucidates this category of risk as being directly associated with general-purpose models. Under Article 51 of the EU AI Act, a general-purpose AI model is deemed to possess systemic risk if it satisfies one of two conditions: (a) it demonstrates high-impact capabilities, assessed using appropriate technical tools, benchmarks, and indicators; or (b) the European Commission determines, either on its own initiative or following a qualified alert from the scientific panel, that the model has equivalent capabilities or impact, based on criteria in Annex XIII. A presumption of high-impact capability arises when the cumulative computational power employed during the model's training surpasses 10^25 floating-point operations (FLOPs); this quantitative threshold serves as a proxy for the potential magnitude of the model's societal and systemic influence[6]. The definition highlights two central features – namely, a *large-scale* and *significant societal impact* that extends beyond individual harm to affect *broader segments of the society*. Thus, the concept of systemic risk cannot be reduced solely to potential damage to "safety, security, and fundamental rights", as frequently invoked within the AI Act; rather, it encompasses these domains in their collective and structural dimensions. Recital 110 further clarifies the nature of

[6] Not all general-purpose models, therefore, entail systemic risk. In the absence of such risk, they are subject primarily to obligations concerning documentation and compliance with copyright requirements. When systemic risk is present, however, additional obligations apply.

systemic risk by providing a more detailed enumeration of its various possible manifestations. This definition of systemic risk aligns with those found in the literature, particularly with the work of Kaufman and Scott (2003: 372): «Systemic risk refers to the risk or probability of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, and is evidenced by co-movements (correlation) among most or all parts». Other characteristics, such as complexity and the cascade effect, are elaborated in the recently published *General-Purpose AI (GPAI) Code of Practice*, which offers a clearer framework for interpreting this section of the AI Act. This transitional document seeks to ensure the presumption of conformity, specify how providers can fulfil the obligations set out in Articles 53-55 of the AIA, maintain up-to-date technical documentation, and support the continuous assessment and mitigation of systemic risks. In particular, Appendix 1 of Chapter 3 contributes to the classification and identification of various types of systemic risks, specifying their nature and sources.

Systemic risks are here primarily characterised by their high-impact capabilities, as defined in Articles 3(64)-(65) of the AI Act, their potential to exert a significant influence on the Union market, and their capacity to propagate widely across the value chain. These elements make such risks particularly complex and far-reaching. Contributing factors further intensify this dynamic. The level of risk tends to increase in proportion to the model's capabilities and diffusion, while the speed at which these risks can materialise is often remarkably high. Moreover, systemic risks in AI may trigger cascading effects that are difficult, if not impossible, to reverse. Their impact is frequently asymmetric, meaning that the actions of a few actors – or even isolated events – can generate disproportionately large and potentially disruptive consequences.

Four main categories of systemic risk emerge: 1) Chemical, biological, radiological and nuclear (CBRN) the facilitation of attacks involving chemical, biological, radiological, or nuclear weapons; 2) Loss of control, models that evade human oversight or enable the autonomous research and development of AI; 3) Cyber offence, offensive capabilities that enable large-scale, sophisticated cyberattacks; 4) Harmful manipulation interference in decision-making processes that threatens fundamental rights and democratic values.

In light of this framework, we can now turn to some reflections on the notion of risk, systemic risk and, more specifically, on their relationship.

As previously noted, the European legislator provides a specific definition of «risk» as «the combination of the probability of an occurrence of harm and the severity of that harm». This conception emphasizes a realistic and

measurable notion of risk, typical of technical domains where evaluation is expected to rely on objective system properties. Yet, such an approach faces several challenges. Abstract dimensions such as human dignity or personal integrity resist objective calculation (Chaberlain, 2023). Moreover, the continuously evolving nature of AI systems – still undergoing rapid expansion – further undermines the feasibility of approaches grounded in predictability and quantitative assessment, calling for broader, more adaptive understandings of risk. As Mahler states:

> Such calculations work best under the assumption that the future conditions of the relevant context are comparable to past conditions, which may be questionable when AI system continues to learn and evolve, for example. If the future is different from the past, calculations may become problematic (Mahler, 2021: 260).

This framework reveals a gap between the definition of risk proposed in the document and the actual nature of the threats posed by contemporary AI systems. On the one hand, the document reproduces objective and realistic interpretations aimed at the technical and calculable resolution of risk; on the other hand, such strategies conflict with the incalculable, unpredictable, and autonomous nature that characterizes AI. It seems that, in the words of Kaminski, «the choice to use risk regulation reflects a particular epistemology: the notion that such AI systems are just math, uncovering some ground truth then contingent social facts» (Kaminski, 2023: 1400).

Moreover, the definition of risk appears to diverge from the way risk is described and assessed. As Novelli observes, risk identification relies on predetermined categories shaped by the values of the European Union. This approach, however, risks being overly static and may underestimate alternative sources of risk arising from different uses – for instance, video games.

The introduction of the category of systemic risk appears to address this concern, as it is characterized by greater impact, heightened uncertainty, and increased complexity. However, as indicated in Definition 65, systemic risk is confined exclusively to GPAI models. This restriction limits the systemic dimension to a narrow subset of models – a position that is, at the very least, debatable, given that in a digital risk society most risks exhibit systemic characteristics. It is therefore necessary to clarify the relationship between the two terms: are they distinct and mutually exclusive categories, or can they overlap?

To address this question, two aspects must be considered.

First, the two levels concern different objects: on the one hand, AI systems, and on the other, AI models. A model is a specific program trained on

data to perform a defined task, such as image recognition or text translation. A system, by contrast, is a broader ensemble of components that includes one or more AI models, together with software, hardware, and data, in order to solve a more complex end-to-end problem. This implies that a system may encompass a model within it.

The second aspect concerns the identification of risk. In the case of systems, the classification of risk levels is based on use, following the so-called "risk pyramid". In contrast, general-purpose AI models are not differentiated by use but rather by the model's computational power. These two dimensions reveal a discontinuity between the risk pyramid and the notion of systemic risk.

Overlaps between the two categorizations may therefore occur. AI systems classified as presenting unacceptable, high, low, or minimal risk may, in fact, be based on the application of GPAI models. At the same time, one might also ask whether systems that do not rely on GPAI models could nonetheless pose a systemic risk. This latter possibility appears to be excluded, as definition 65 makes explicit reference to GPAI models. However, it seems plausible that systemic risks may arise more frequently than anticipated by the law. As previously noted, the current landscape of the digital risk society generates increasing levels of uncertainty and unpredictability – potentially on a global scale – thereby challenging the boundaries set by the existing regulatory framework.

The European regulatory framework, while representing a pivotal step in the global governance of artificial intelligence, remains grounded in a predictive and calculable logic that struggles to grasp the complexity, fluidity, and inherent uncertainty of today's digital ecosystem. Such an approach reflects a primarily technical conception of risk – one that can ostensibly be mitigated through interventions at the system level rather than through consideration of the social equilibria in which such systems are embedded. The rationale behind this orientation is clear: the AI Act is modelled on product safety regulation, upon which the protection of fundamental rights is superimposed. Consequently, the focus is directed more toward the product itself than toward the individuals exposed to potential harm. Yet, this orientation calls for a further reflection – one that takes into account not only the notion of risk but also that of the subject at risk. Given the unpredictable nature of these emerging threats, concentrating solely on the product fails to address the underlying conditions that make harm possible: the exposure and vulnerability of individuals (Zanotti *et al.*, 2024). This limitation underscores the need for a more adaptive approach – one capable of integrating

technical, social, and ethical dimensions into the governance of technological risk. Introducing the notion of vulnerability thus broadens the concept of risk, shifting attention from the systems that generate it to the contexts and subjects who endure it. A vigilant observation of the conditions that enable injustice, discrimination, manipulation, and restrictions of freedom requires, above all, a gaze directed toward the human before the machine.

6. *Conclusions*

The historical and conceptual evolution of risk – from antiquity to the contemporary digital era – reveals a profound transformation: from a calculable, probability-based notion to an uncertain, global, and systemic phenomenon. In the context of AI, risks are not only technical but sociotechnical, affecting individuals and societies through interconnections, opacity, and potential cascading effects. Systemic risk, particularly in general-purpose AI models, highlights how localized failures or autonomous system behavior can propagate across social, economic, and technological networks.

The EU AI Act represents a pioneering step in risk-based regulation, yet its emphasis on calculable, product-centered risk contrasts with the inherent unpredictability and systemic nature of AI. This tension underscores the need for an integrated approach that combines technological safeguards, regulatory oversight, and ethical reflection, shifting the focus from systems alone to the vulnerabilities of the individuals and contexts they affect. Ultimately, contemporary digital risk requires a governance framework that goes beyond technical mitigation, addressing social, ethical, and structural dimensions of harm.

References

Allen, F., Carletti, E. (2013). *What Is Systemic Risk?*, in «Journal of Money Credit and Banking», 45 (1), pp. 121-127.

Anders, G. (1962). *Burning Conscience. The case of the Hiroshima pilot, Claude Eatherly, told in his letters to Gunther Anders, with a postscript for American readers by Anders*, Monthly Review Press, New York.

Angwin, J. *et al.* (2023), *Machine bias. There's software used across the country to predict future criminals. And it's biased against blacks*, in «Propublica», May 23, https://www.propublica.org/article/machine–bias–risk assessments–in–criminal–sentencing.

Aven, T., Renn, O. (2019). *Some foundational issues related to risk governance and different types of risks*, in «Journal of Risk Research», 1, pp. 1-14.

Beck, U. (1992). *Risk Society. Towards a New Modernity*, SAGE, London.

Beck, U. (2008). *Risk Society's 'Cosmopolitan Moment*, in «ComCiência» [online], n. 104.

Bernstein, P.L. (1996). *Against the Gods. The Remarkable Story of Risk*, Wiley, New York.

Bostrom, N. (2014). *Superintelligence. Paths, Dangers, Strategies,* Oxford University Press, Oxford.

Bradford, A. (2020). *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, New York.

Cappelen, H. *et al.* (2025). *AI Survival Stories: a Taxonomic Analysis of AI Existential Risk*, in «Philosophy of AI», 1, pp. 1-19.

Casonato, C., Olivato, G. (2024). *AI Regulation in Europe: Exploring the Artificial Intelligence Act*, in A. Fabris, S. Belardinelli (eds), *Digital Environments and Human Relations. Human Perspectives in Health Sciences and Technology*, Cham, Springer.

Chamberlain, J. (2023). *The Risk-Based Approach of the European Union's Proposed Ar-tificial Intelligence Regulation: Some Comments from Tort Law Perspective*, in «Euro-pean Journal of Risk Regulation», 14, pp. 1-13.

Cinelli, M. *et al.* (2020). *The echo chamber effect on social media*, in «PNAS», 118, 9, e2023301118.

Coeckelbergh, M. (2015a). *Human Being @ Risk: Enhancement, Technology, and the Evaluation of Vulnerability Transformations*, Springer, Cham.

Coeckelbergh, M. (2015b). *The tragedy of the master: automation, vulnerability, and distance*, in «Ethics and Information Technology», 17, pp. 219-229.

Covello, V.T., Mumpower, J. (1985). *Risk Analysis and Risk Management: An Historical Perspective*, in «Risk Analysis», 5, p. 108.

Dean, M. (1998). *Risk, Calcuable and Incalculable*, in « Soziale Welt», 49, pp. 25-42.

Douglas, M. (1992). *Risk and Blame. Essays on Cultural Theory*, Routledge, London.

Fabris, A. (2018). *Ethics of Information and Communication Technologies*, Springer, Cham.

Fabris, A. *et al.* (2024). *Towards a Relational Ethics in AI. The Problem of Agency, the Search for Common Principles, the Pairing of Human and Artificial Agents*, in A. Fabris, S. Belardinelli (eds), *Digital Environments and Human Relations. Human Perspectives in Health Sciences and Technology*, vol. 150, Springer, Cham.

Fraser, H., Bello y Villarino, J.M. (2023). *Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough*, in «European Journal of Risk Regulation», pp. 1-16.

Giddens, A. (2000). *Il mondo che cambia. Come la globalizzazione ridisegna la nostra vita,* il Mulino, Bologna.

Giusti, S., Piras, E. (2020). *Democracy and Fake News. Information Manipulation and Post-Truth Politics*, Routledge, London.

Guan, H. *et al.* (2022). *Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making*, in «Behavioral Sciences» 12, 9, p. 343.

Guerrero *et al.* (2018). *Analysis of Racial/Ethnic Representation in Select Basic and Applied Cancer Research Studies*, in «Scientific Reports», 1, pp. 1-8.

Hagendorff, T. (2022). *Blind spots in AI ethics*, in «AI Ethics», 2, pp. 851-867.

Hendrycks, D., Mazeika, M., Woodside, T. (2023). *An Overview of Catastrophic AI Risks*, in «Arxiv», https://doi.org/10.48550/arXiv.2306.12001.

Hinds, J. *et al.* (2020). *"It wouldn't happen to me": privacy concerns and perspectives following the Cambridge Analytica scandal*, in «International Journal of Human-Computer Science», 143, 102498.

IRGC (2018). *Guidelines for the Governance of Systemic Risks*, International Risk Governance Center (IRGC), Lausanne.

Kaminski, M.E. (2023). *Regulating the Risks of AI,* in «Boston University Law Review», 103, pp. 1347-1411.

Kaufman, G., Scott, K. (2003). *What Is Systemic Risk, and Do Bank Regulators Retard or Contribute to It?*, in «Independent Review», 7, pp. 1-31.

Luhmann, N. (1992). *Risk: a sociological theory*, de Gruyter, New York.

Lupton, D. (1999). *Risk*, Routledge, London.

Lupton, D. (2016). *The diverse domains of quantified selves: self-tracking modes and dataveillance*, in «Economy and Society», 45, 1, pp. 101-122.

Mahler, T. (2021). *Between risk management and proportionality: the risk-based approach in the EU's Artificial Intelligence Act Proposal*, in «Nordic Yearbook of Law and Informatics», pp. 245-267.

O'Neill, B. (2018). *The great Cambridge Analytica conspiracy theory*, in «The Spectator», 21 March 2018 (https://www.spectator.co.uk/article/the–great–cambridge–analyt-ica–conspiracy–theory/).

OECD (2003). *Emerging Risks in the 21st Century. An Agenta for Action*, OECD Publications Service, Paris.

Parisier, E. (2020). *The Filter Bubble: What the Internet Is Hiding from You*, Penguin, London.

Poledna, S., Rovenskaya, E., Dieckmann, U., Hochrainer-Stigler, S., Linkov, I. (2020). *Systemic risk emerging from interconnections: the case of financial systems,* in W. Hynes, M. Lees, J. Müller (eds), *Systemic thinking for policy making: the potential of systems analysis for addressing global policy challenges in the 21st century*, OECD Publishing, Paris.

Renn, O. *et al.* (2022). *Systemic Risks from Different Perspectives*, in «Risk Analysis», 42, 9, pp. 1902-1920.

Sillman, J. *et al.* (2022). *Briefing Note Systemic Risks: Review and Opportunities for Research, Policy and Practice from the Perspective of Climate, Environmental and Disaster Risk Science and Management*, International Science Council, Paris.

Sundberg, L. (2024). *Towards the Digital Risk Society: A Review*, in «Human Affairs», 34, 1, pp. 151-164.

Terenzio, F. (2025). *Systemic Vulnerability: From AI Systems to Environmental Systems*, in «Topoi».

Trump, B.D. *et al.* (2021). *Multi-Disciplinary Perspectives on Systemic Risk and Resilience in the Time of COVID-19,* in I. Linkov *et al.* (eds), *COVID-19: Systemic Risk and Resilience, Risk, Systems and Decisions*, Springer, Cham.

Van de Poel (2016). *An Ethical Framework for Evaluating Experimental Technology*, in «Sci Eng Ethics», 22, pp. 667-686.

Zanotti, G., Chiffi, D., Schiaffonati, V. (2023). *AI-Related Risk. An Epistemological Approach*, in «Philos. Technol.», 37, p. 66.

Abstract

This article explores the historical evolution of the concept of risk and its contemporary applications in the domain of artificial intelligence. Building on Beck's notions of "global risk" and "risk society", we argue that today's context can be described as a digital risk society, *increasingly marked by uncertainty and unpredictability. The idea of systemic risk – widely discussed in complexity science and economics – emerges as a particularly suitable conceptual tool for interpreting this scenario. The centrality of risk is further evidenced by the EU's AI Act, which adopts a risk-based approach. We examine this regulation to analyze the interaction between risk and systemic risk, highlighting the limitations of this framework.*

Keywords: risk; systemic risk; AI Act; vulnerability.

Silvia Dadà
Università di Pisa
silvia.dada@unipi.it