XV Convegno annuale
dell'Associazione Italiana dei Professori Universitari
di Diritto Commerciale "Orizzonti del Diritto Commerciale"

"Impresa e mercati: numeri e computer science"
Roma, 23-24 febbraio 2024

Michele Mozzarelli

Professore Ordinario di Diritto Commerciale,
Università Cattolica del Sacro Cuore di Milano

Giulia Schneider

Ricercatrice di Diritto dell'Economia,
Università Cattolica del Sacro Cuore di Milano

AI and Strategic Decisions: Facing the Incompleteness

1. *Introduction*

Is it possible to entrust an algorithm with the definition of a company's industrial policy, such as a significant decision like a merger or the identification of members for an optimal composition of a board of directors?

The answer to this question appears increasingly disconnected from the technological sphere, especially since the advent of artificial intelligence (AI), whose impact is currently being pervasive in commercial activities and beyond.

The distinctive features of AI systems lie in their ability to achieve a given set of objectives autonomously, thereby generating predictions, recommendations, and decisions. The significant revolution brought about by artificial intelligence is to be precisely related to its expanding scope of application, increasingly covering the entire decision-making process, thereby leading to optimisation and efficiency gains.

The source of concern in entrusting a significant decision to an algorithm is not the technical capacity of the algorithm to make decisions but the legal risks associated with such a capability. As widely recognised, optimising algorithmic decision-making comes with inherent risks.

European regulators have variously addressed the risks stemming from the adoption of new technologies by businesses. Regulations such as the GDPR, the proposed AI Act, the DSA, DMA, and, lately, the DORA for the financial sector have introduced different tools to control some of the risks associated with these new technologies. This continuously evolving normative framework within the EU aims to guide the efficiency-driven advancement of digital transformation in both the private and public sectors, along the lines of consolidating legitimacy paradigms.

In particular, our understanding is that the chosen approach to maintain an anthropocentric conception of ongoing digital transformation focuses at the procedural level on the 'human in the loop' approach. As elucidated further in subsequent sections, AI—broadly understood as automated systems—is treated by the regulations above as a product, the safety of which, in terms of preventing harm to third parties, must be ensured through human monitoring. Little attention is paid to the whole decision-making process,

including the human decision-making component and its effective interaction with employed AI systems.

Against this backdrop, this study interrogates the role of AI systems—and related legal requirements—in strategic corporate governance decisions, characterised by a high degree of incompleteness and, consequently, discretion. It illustrates the inadequacy of current European regulatory developments in the IT field, primarily based on a product liability approach, to govern the complexities underlying the introduction of AI tools in directors' strategic choices.

Given the core elements of corporate governance and its specific accountability needs, the study demonstrates how algorithmic decision-making cannot be an exclusive strategy for strategic decisions characterised by unpredictability and incompleteness.

Therefore, the ultimate objective of the analysis is to underscore the persisting gap in both legislation and scholarly literature regarding the need to shift from a product-based and risk-minimization approach in AI-driven decision-making to an integrated approach wherein the decision-making process is shaped by the complex, yet balanced, interaction between two distinct decision-making components, namely the human and the algorithmic decision-maker. This presents an ambitious, albeit necessary, challenge.

In line with the proposed conception concerning strategic and high-impact decisions, legal strategies should not only focus on the machine, which outcomes are supervised by humans, but also on the interaction between humans and machines.

To this end, the paper is structured as follows: the first section elucidates the regulatory perspective on AI accountability in accordance with the 'human in the loop' and thus a product-based approach.

3

The second section contextualises the presently dominant approach of humans in the loop as a risk-mitigating factor for potential harms arising from deployed AI models in the Corptech domain. It illustrates how the polarisation characterising corporate scholarship around Corptech solutions can be viewed as a direct by-product of the endeavour to emphasise the distinctions, also in terms of AI implications, between corporate obligated procedural (protocol-based) decisions and discretionary decisions. Hence, in the third section, we contend that in corporate strategic decisions, the human-in-the-loop approach is inadequate due to the combination of the actual incompatibility of AI and corporate governance legal strategies and the risk of overload of the human controller. In the fourth and final section, we propose a change of perspective in AI regulation aimed at retaining humans in command and focusing on the interaction between humans and machines rather than solely on the algorithms. To this end, we advocate for a stronger emphasis on human enhancement, algorithm ergonomics, and a legal strategy-driven design of AI. Conclusions ensue.

SECTION I – AI as a Product

   2. *The State of the Art of EU Digital Regulations*

The employment of AI tools for decision-making purposes has raised some non-neglectable practical problems, directly deriving from the specific technological features of AI systems. The most common, and widely documented, concerns are related to the perpetuation of various type of biases (historical, social, statistical ones)[1], and the structural obscurity of AI

---

* Although this paper is the result of joint work, Sections I and II should be attributed to Giulia Schneider and Sections III and IV to Michele Mozzarelli.

models[2]. These features have resulted into explainability[3] and understandability[4] gaps of AI-driven decisions, which have soon climbed up to the forefront of the European regulatory agenda. Indeed, the focus of emerging IT regulations has been majorly set onto the objective of achieving transparency of automated technologies used for decision-making, as a means to control the threats to the protection of fundamental rights in accordance with a risk-based approach.

Overall, the European IT framework is meant to control the risks directly arising from the impact of the employment of technological tools on the protection of personal fundamental rights and adversely affecting end-users as the addressees of automated decision-making processes; the businesses making use of these technological tools, amplifying in particular their legal and operational risks; and, finally, also the broader economic system, in terms of market integrity and stability[5].

This risk-based approach targeting the protection of the personal fundamental rights enshrined in the European Charter is one of the

---

[1] *Ex multis* S. BAROCAS — A.D. SELBST, *Big data's disparate impact*, in *California law review* 2016, 104, 3, 694-712; P.T. KIM, *Data-driven discrimination at work*, in *Wm. & Mary L. Rev.* 2016, 58, 883-892.

[2] S. ZUBOFF, *The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier*, Public Affairs, New York (2019), passim.

[3] See S. ALI ET AL., *Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence*, in *Information Fusion* 2023, 99, 101805. See also S. WACHTER — B. MITTELSTADT — C. RUSSELL, *Counterfactual explanations without opening the black box: Automated decisions and the GDPR*, in *Harv. JL & Tech.* 2017, 31, 849-853.

[4] A.D. SELBST — J. POWLES, *Meaningful information and the right to explanation*, in *International Data Privacy Law* 2017, 7, 4, 239-242; G. MALGIERI — G. COMANDÉ, *Why a right to legibility of automated decision-making exists in the general data protection regulation*, in *International Data Privacy Law* 2017, 7, 4, 252-256.

[5] In respect to the consideration of economic interests affected by automated tools, see Regulatory Obstacles to Financial Innovation Experts Group-ROFIEG, *30 Recommendations on Regulation, Innovation and Finance* (13 December 2019) https://finance.ec.ef.uropa.eu/publications/final-report-expert-group-regulatory-obstacles-financial-innovation-30-recommendations-regulation_en.

distinctive features of the emerging European "digital constitutionalism"[6], as opposed to the American and Chinese models of technology regulation[7]. Although with substantial differences, the IT regulations so far issued at European level share the common feature of focusing on the structural features of technologies enabling the pursuing of accountability objectives[8]. At general level, accountability of technological means relates to the ability to establish whether a technological system works in conformity with substantive and procedural standards. Where these standards are not accomplished, controllers' liability arises, as a result of the failed task to control the various risks stemming from the departure from applicable standards[9]. One of the primary means to hold an automated system accountable, is the establishment of transparency requirements.

First steps regarding automated decision-making accountability have been taken by the General Data Protection Regulation (GDPR)[10] which in its turn from a data protection to a data management regulatory paradigm[11] has obliged businesses processing personal data to ensure that the processing operations they were performing were externally traceable and justifiable[12], through the production of "meaningful explanations" for the ultimate purpose of protecting data subjects' (personal) fundamental right to data protection.

---

[6] G. DE GREGORIO — P. DUNN, *The European risk-based approaches: Connecting constitutional dots in the digital age*, in *Common Market Law Review* 2022, 59, 2, 479-483.

[7] A. BRADFORD, *Digital empires: The global battle to regulate technology*, Oxford, Oxford University Press, 2023, *passim*.

[8] I. CARNAT, *Ethics Lost in Translation: Trustworthy AI from Governance to Regulation*, in *Opinio Juris in Comparatione* 2023, 4, 104-110.

[9] G SCHNEIDER, *Accountability*, in G. Comandé (ed.), *Encyclopaedia of Law and data science*, Edward Elgar, 2022, 7 ff., 7.

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

[11] For a recent analysis on this see P. BALBONI — K. FRANCIS, *Data protection as a corporate social responsibility*, Cheltenham, Edward Elgar Publishing, 2023, in particular at 22-47.

[12] See principles of transparency and accountability under art. 5 GDPR.

Very similarly to the GDPR, also the proposed AI Act[13], defines AI transparency from a functional standpoint, and links this structural feature to the achievement of "compliance with the relevant obligations of the user and of the provider" set out in the Regulation for those systems posing a high-risk to personal fundamental rights[14].

In the (difficult) effort of regulating the highly changing landscape of AI systems, the proposal has opted for establishing only programmatic objectives for those systems that fall under the category of high-risk systems[15].

Interestingly, in accordance with the AI Act's objective of protecting users' personal fundamental rights, the systems listed as high risk under Annex III of the proposal do not encompass many applications of AI in the business sector.

2.1 *The EU Regulatory Technique: Human Subjects as Monitoring Gatekeepers*

From a functional standpoint, the recalled EU IT transparency-based framework requires that the systems employed for autonomous decision making have some structural features which enable the activation of a human-machine interaction. Enacted rules regarding employed IT systems indeed place first limits to initiated processes of de-humanization of decision-making triggered by optimization promises, in accordance with an anthropocentric conception of the ongoing digital transformation, which is based, at procedural/methodological level, on the 'human in the loop' command. The resulting regulatory paradigm sets the focus on the

---

[13] The last version of the AI Act has been agreed on the 26th January 2024. Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Analysis of the final compromise text with a view to agreement, 26 January 2024, 5662/24 (hereafter AI Act).
[14] See AI Act, art. 13.
[15] See annex III of the proposed AI Act.

functional control relationship between human and machine: in respect to the automated decision-making loop, human subjects are needed to control and manage risks associated with the technological infrastructure.

This standpoint is clearly expressed by recital 1 of the AI Act, which recalls a paradigm of human-centric and trustworthy AI. Recital 4a of the AI Act confirms this stance, highlighting that "as a pre-requisite, artificial intelligence should be a human-centric technology", which should "serve as a tool for people, with the ultimate aim of increasing well-being". The human-centric approach to AI is thus declared to be at the very core of the European legislative intervention in the field of AI. Nonetheless, this European manifesto is only partly addressed in the provisions the proposed AI Act lays down. The core problem of retaining AI-driven decisions under the command of human subjects remains largely unaddressed, as if it was taken for granted by European regulators. As shaped by normative requirements, indeed, the role of human subjects in an AI-driven decision-making process is that of supervising and, in accordance with the risk-based approach adopted both by the GDPR and the AI Act, of minimising the possible risks attached to entirely technological-driven outputs.

Ultimately, indeed, both transparency and thus accountability requirements are meant to enact a human scrutiny over the automated decision-making loop so as to preserve the fairness of the processing, majorly given by the absence of biases affecting the automated decision-making loop to the detriment of the vulnerable parties to the [16].

An example of technology-related accountability requirement under the GDPR is given by the supervisory obligations attributed to data controller and processors regarding their processing means. Art 35 GDPR, for example, imposes onto businesses processing personal data the burden of

---

[16] G SCHNEIDER, *Fairness*, in G. COMANDÉ (ed.), *Encyclopaedia of Law and data science*, Edward Elgar, 2022, 168 ff., 171-173.

measuring the threats to data subjects' fundamental rights arising from enacted processing activities, in the form of a Data Protection Impact Assessments (DPIA)[17].

With specific regard to automated decision-making processes, art. 22 GDPR, establishes a right not to be subject "to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her". The provision has been recently interpreted by the CJEU in the Schufa Judgment as an outright prohibition of fully automated decisions, echoing an interpretation given by a strand of the literature.[21]

This prohibition is to be first interpreted as banning automated decisions, unless human monitoring activities over the system's features, eventual biases and ultimate risks are properly fulfilled. Along these lines, the same art. 22 GDPR establishes a "right to intervention" of the user, thus a right to challenge machine-generated outputs, which is strictly related to the right to object to algorithmic determinations.

Along similar regulatory lines, the last version of the AI Act requires high-risk systems to be subject to human oversight. This supervisory function, however, as revealed by the analysis of the same provision under art. 14 AI Act, is of mere monitoring support of the output of the automated decision and not much of the algorithmic process leading to the decision.

Indeed, despite the first paragraph demands high-risk AI systems to be designed and developed in such a way, including with appropriate human-machine interface tools, "that they be effectively overseen by natural persons during the period in which the AI system is in use" , the following paragraph 2 admits that this human oversight shall be ensured so as to prevent or minimise the "risks to health, safety or fundamental rights that

---

[17] M.E. KAMINSKI — G. MALGIERI, *Algorithmic impact assessments under the GDPR: producing multi-layered explanations*, in *International Data Privacy Law* 2021, 11, 2, 129-131.

may emerge when a high-risk system is used in accordance with its intended purpose or under conditions of reasonable foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter". As this wording suggests thus, human oversight shall be functional just to detect eventual risks to fundamental rights, but it is not meant to touch on or interfere with a decision- making process that could thus also be of fully automated nature[18].

If it doesn't regard directly the algorithmic process, human oversight cannot but refer on the output of the (also fully automated) decision, as confirmed by the same article 14(4) requiring the human overseer to "correctly interpret the high-risk system's output" (lett. c); as well as "to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system" (lett. d).

In accordance with the oversight functions under art. 14 AI Act, the human subject can either decide to "intervene on the operation of the high-risk system", mainly through interrupting the system "though a "stop" button or a similar procedure" (lett. e) (thus, refusing to use the output of the automated system), or internalize such output in the final decision outcome with a mere warning to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias') (lett. b).

The regulatory perspective is thus the one of an AI-driven decision-making "loop" in which the human component serves as an external watchdog of the technological decision-making outcome and of associated risks to

---

[18] This perspective appears to be further confirmed by the erasure in the very last circulated version of the AI Act of the overarching principle of non-exclusivity of AI decisions under a previously proposed art. 4a AI Act, whose introduction had been suggested by the European Parliament as a direct expression of the human-centric approach to AI expressed in recitals 1 and 4 of the AI Act.

health, safety or fundamental rights recalled under art. 14(2) AI Act ("human in the loop" approach). In other words, AI– broadly intended as automated systems– is treated by the mentioned regulations as a product, the safety of which– in terms of prevention of the causation of harms onto third parties– must be ensured by human monitoring and evaluation. Nothing is said about the decision-making process considered as a whole, as given, eventually, also by the human decision-making component.

It is no coincidence that recent legal literature[19] has identified various critical issues concerning the human oversight remedy placed at the heart of the European regulation on artificial intelligence. Among these, it should first of all be recalled that the monitoring duties looked at by the European legislator are majorly enabled by the initial design options of the machine as determined by the producer.

Indeed, in the very end, the business user is forced to circumscribe its monitoring capabilities within the instructions given by the manufacturer.

The problem of the human chain of responsibility, from the manufacturer to the user of the system, has been variously commented on by studies in the field of torts[20].

The complexity of the human network - and thus of responsibility - has been posited by some as the basis of a necessarily complex liability system that must accompany the life cycle of an automated system. However, as suggested by the proposed framework on AI liability, lastly proposed by the European Commission in September 2022, and given by a proposal for a revision of the directive on product liability and a proposal for a directive

---

[19] See in particular, R. CROOTOF — M.E. KAMINSKI — W.N. PRICE II, *Humans in the Loop*, in *Vanderbilt Law Review* 2023, 76, 2, 503-505.

[20] G. COMANDÉ, *Multilayered (Accountable) Liability for Artificial Intelligence*, in S. LOHSSE — R. SCHULZE — D. STAUDENMAYER (a c. di), *Liability for Artificial Intelligence and the Internet of Things*, Nomos Verlagsgesellschaft mbH & Co. KG, 2019, *179-180.*

on AI liability[21], the complexity of a possible liability system regarding AI risks has been ultimately simplified through the positioning of the ultimate liability burden onto the provider of a given AI system. As opposed by the initial stance initially proposed by the European Parliament[22], the proposal of a revision of the Product liability directive[23] tackles only the liability of manufacturers, governed by a strict liability regime. Also the proposed AI liability directive[24] only narrowly focuses on the liability of manufacturers and users in case of damages arising from high risk systems under the AI Act, without touching upon the more complex issue of the user's (in our case the business') extra-contractual (fault-based) liability of low-risk AI systems under the AI-related regulatory framework. Hence, the envisaged liability regime for harms arising from AI systems confirms the product-centred viewpoint taken by the European regulator in the matters of AI. It thus ensures the coverage of torts arising from AI systems identifying the producers as the main source of liability and the primary spot of deterrence[25].

---

[21] EUROPEAN COMMISSION, *Liability Rules for Artificial Intelligence*, https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en

[22] EUROPEAN PARLIAMENT, *Civil Liability Regime for Artificial Intelligence*, September 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654178/EPRS_STU(2020)654178_EN.pdf. ID., *European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence*, 20 October 2020, (2020/2014(INL)).

[23] EUROPEAN COMMISSION, *Proposal for a Directive of the European Parliament and of the Council on liability for defective products*, COM/2022/495 final.

[24] EUROPEAN COMMISSION, *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, COM/2022/496 final. See more EUROPEAN PARLIAMENT, *Artificial intelligence liability directive*, February 2023, https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf.

[25] G. WAGNER, *Liability Rules for the Digital Age: – Aiming for the Brussels Effect –*, in *Journal of European Tort Law* 2023, 13, 3, 191 ff., in particular at 196-197. As has been suggested by this literature, the option of leaving the liability of users of AI unaddressed (at least with specific tort rules) reflects the European regulators' intention not to excessively freeze the market of AI systems and thus their use by businesses. However, the implicit appeal to the more general technology neutral liability rules for fault for regulating users' liability raises a series of unanswered questions regarding how to tailor the general precepts of extra-

2.2 *The Last Generation of Human in the Loop Rules: DORA*

In line with the above identified approach, also financial sectoral regulations have adopted human in the loop-styled requirements. This is well reflected in the Regulation on the digital operational resilience of the financial sector[26], which builds on the technological resilience requirements first traced in the Mifid II framework[27] for the protection of fundamental economic interests of financial stability and market integrity[28].

Referring to the governance of ICT risks in the financial sector, the DORA appears to be a last-generation of "human in the loop" regulation, which lays down external limits to the employment of new technologies, if not properly supervised through a complex (human) organization.

Similarly to what occurs under the above recalled general IT regulations, also the DORA regulation remains loyal to a product-based approach to technologies: it indeed calls for the adaptation of financial entities' corporate governance structures to the early prevention, detection and cure of ICT risks. The risks that are targeted by this piece of regulation are those potentially impacting the financial institutions' digital operational resilience, digital operational resilience, which art. 2(1) DORA declines as

---

contractual liability in respect to users' monitoring duties required by the AI Act: when is a monitoring activity conducted negligently? Which is the monitoring standard that the user has to prove so as to prove its diligence? How should this liability be matched with the manufacturers' one? The Human in the loop paradigm affirmed by the AI Act triggers these queries, and no response is to be found in the neighbouring emerging AI liability framework.

[26] Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011 (hereafter DORA).

[27] Cfr. art. 47 MIDIF II. Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU.

[28] EUROPEAN COMMISSION, *Digital Finance Package*, https://finance.ec.europa.eu/publications/digital-finance-package_en.

the "ability of a financial entity to build, assure and review its operational integrity and reliability by ensuring, either directly or indirectly through the use of services provided by ICT third-party service providers, the full range of ICT-related capabilities needed to address the security of the network and information systems which a financial entity uses, and which support the continued provision of financial services and their quality, including throughout disruption".

Moving beyond a mere fundamental rights-based conception of IT accountability, the DORA thus targets the control of the economic interests potentially affected by the digitalization of financial business structures[29]

In respect to these ICT risks, hence, human subjects, and in particular directors, serve as a liability site, so as to remedy eventual, legally relevant automated-driven "adverse effects". In this perspective, the turning point of the DORA in respect to the GDPR and also the proposed AI Act lies in the fact that the supervision of technologies becomes a corporate governance matter[30].

To these ends, assuming that, in line with strict ICT- regulations, technologies have to be designed in a manner that enables human supervision, the DORA has specified *how* technologies in the financial sector have to be supervised. It has laid down specific procedural and organizational requirements for the management of ICT tools by financial entities, evolving around the principle of the management body's "ultimate responsibility"[31]. The management body, in the continuous interaction with

---

[29] ICT risks are largely defined as "any reasonably identifiable circumstance in relation to the use of network and information systems which, if materialised, may compromise the security of the network and information systems, of any technology dependent tool or process, of operations and processes, or of the provision of services by producing adverse effects in the digital or physical environment". Art. 3(5) DORA.

[30] D.A. ZETZSCHE ET AL., *Artificial Intelligence in Finance: Putting the Human in the Loop,* in *Sydney Law Review* 2020, 43 ff., 72-73.

[31] Recital 45 DORA.

ICT and risk management function[32], would need to take a "pivotal and active role in steering and adapting the ICT risk management framework and the overall digital operational resilience strategy" [33]. In this way, it requires financial entities' directors to take a result-oriented and functional approach that requires a closed monitoring to employed technologies, which needs to be sensitive to the mitigation of ICT risks also beyond the dictates of specific ICT requirements[34].

Differently from the GDPR and the AI Act, the DORA places a greater focus on the component of human behavior vis à vis technological tools through the lenses of corporate governance structures. However, the methodological approach to the regulation of (financial) businesses' digital assets, be it personal data (under the GDPR); AI systems (under the AI Act) or ICT infrastructures (as addressed by DORA), remains, also under the DORA, that of entrusting to human subjects the oversight and control of the functioning and of the final output of employed technologies.

Under these premises, it can be concluded that the objectives pursued by contemporary IT regulations is the understanding and control by humans of the risks to either personal or economic interests arising from the employment of automated systems. In respect to these risks, human handling automated tools have the function of preventing the occurrence of harms and to monitor expected risks through available regulatory tools (in terms of transparency, explainability, auditing) so as to pull the emergency brake, when needed, and either remove or substitute a faulty device.

However, the monitoring role attributed to the human "controller" and the remedies it provides to him, implicitly suggest that the legislator admits the possibility of an entirely automated decision-making process, provided that

---

[32] Art. 5 DORA.

[33] Recital 45 DORA.

[34] Advocating for a functional role of board's oversight of businesses' compliance choices, J. ARMOUR — J. GORDON — G. MIN, *Taking compliance seriously*, in *Yale J. on Reg.* 2020, 37, 31-39.

the human actor guards this same process. In other words, the regulations issued so far appear to focus on use cases, where the option of a 100% algorithmic decision-making process may be admitted, upon the condition of the enactment of a human in the loop defense. From this specific standpoint, it seems that the technology-driven decision-making processes encompassed by the considered regulatory framework implicitly legitimize AI's replacement of human decision makers.

As has been shown above, the IT regulatory layer demands, in accordance with the human in the loop approach, that employed technologies need to be supervised, thus placing human supervision as a general *external* limit of employed technologies but indirectly allowing, upon conformity with risk minimization and monitoring requirements, the enactment of fully automated decisions.

While the human monitoring paradigm inflecting European IT regulations is to be surely welcomed as a first direct response by the European regulator for the curbing of fast-changing and high-intensive innovation scenarios, it offers just "one view of the cathedral"[35] regarding the implication of employing AI in decision-making process. A slighter bigger piece of our AI-related regulatory cathedral can start to be glimpsed if one tests the described approach against the backdrop of the distinction between organizational or routine business decisions and strategic governance decisions in the corporate realm.

Indeed, in respect to the varied landscape of businesses' decisions, the specific use cases each of the mentioned IT regulations refer to for the purposes of the activation of human-in-the-loop styled requirements (e.g. personal data processing operations under the GDPR; credit or health insurance scoring under the categorization of high-risk systems given by

---

[35] The quote is taken from the famous piece, G. CALABRESI, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85, 6 Harvard Law Review Association (1972) 1089 ff.

Annex III of the AI Act; the general IT infrastructure of financial entities targeted by the DORA), appear to relate to low-discretion administrative structures of contemporary, increasingly digitized corporations.

Certainly, data processing activities addressed by the GDPR or the ICT infrastructure under the DORA Regulation could eventually be employed also in the context of businesses' strategic decisions. However, in light of the above analysis, it can be said the human-in-the-loop requirements have originated in the European IT regulatory framework, in accordance with the legislator's intention, majorly for the purposes of enacting external control checkpoints for decisions regarding the digital transformation of businesses' organizational decisions.

None of the available regulatory tools of the traced framework tackles, to date, the specific risks related to the integration of technologies in the dynamics of corporate governance for the purposes of directors' strategic decisions: no specific and expressed regard is to date given at normative level to the monitoring needs and possible solutions of decisions taken through automated means for the purposes of decisions' involving businesses strategic decisions[36], as the ones regarding fundamental changes.

Interestingly, also the legal debate that has arisen around the phenomenon of the digitization of businesses, shortly known as "Corptech" has missed to consider the (potentially) diverse nature of businesses' automated decisions, and the different governance demands that stem from digitised organizational or strategic corporate structures.

SECTION 2 – The CorpTech Dilemma

---

[36] The distinction between organizational and administrative decisions is recalled also by M. PETRIN, *Corporate Management in the Age of AI*, in *Columbia Business Law Review*, 2019, 3, 966, 983. See also V. KOLBJØRNSRUD — R. AMICO — R. THOMAS, *The promise of artificial intelligence: Redefining management in the workforce of the future*, Accenture Institute for High Performance, 2016.

3. *Setting the Focus: the Corptech Dilemma*

Running in parallel with the evolving regulatory baselines, there is the market reality of business' growing employment of AI models within their organizational and managerial structures.

As increasingly acknowledged by empirical studies at both European[37] and national level[38], autonomous computational models are creeping into the most sensitive corporate decision-making areas. This is well illustrated by the uptake of legal tech tools, predicting for example the conveniency and success rate of a litigation in the corporate realm[39], as well as by the growing employment of AI systems for the evaluation and scenario forecast in the context of mergers[40]. As apparent, litigation or merger decisions are not routine-kind decisions but rather strategic ones.

The application of technological tools to the corporate realm is generally referred to as "Corptech" domain, so as to mirror the developments occurred in the "Fintech" sector.

Exactly because of the growing application of AI-driven decisions in corporate strategic contexts, the literature has started to consider, also

---

[37] EUROPEAN COMMISSION, *Study on the relevance and impact of artificial intelligence for company law and corporate governance*, 2021, https://op.europa.eu/en/publication-detail/-/publication/13e6a212-6181-11ec-9c6c-01aa75ed71a1/language-en.

[38] See for example Assonime, *L'evoluzione dell'organo amministrativo tra sostenibilità e trasformazione digitale Gruppo di lavoro della Giunta Assonime coordinato da Corrado Passera*, 2023/1, https://www.assonime.it/_layouts/15/Assonime.CustomAction/GetPdfToUrl.aspx?PathPdf=https://www.assonime.it/attivita-editoriale/studi/Documents/Note%20e%20Studi%201%20-%202023.pdf, 60-69.

[39] J. ARMOUR- R. PARNHAM- M. SAKO, *Augmented Lawyering*, 1 University of Illinois Law Review (2022) 72 ff., 107-198.

[40] L. LEHOT- E. CHOW, *How Artificial Intelligence is Disrupting the Dealmaking Process*, Medium (18 September 2023) https://lehotlouis.medium.com/how-artificial-intelligence-is-disrupting-the-dealmaking-process-eb6a2ae3bf4c.

under the pressure of evolving practices especially in Asian countries[41], the extreme hypothesis of the substitution of directors by machines, in the form of a robo-director[42], or even of an entirely automated-driven board, as occurs in the case two different hypothesis of a roboboard[43] or of the DAOs[44].

Against the backdrop of these possible– in some cases actual– transformations of the board of directors, the the doctrinal debate appears to have evolved around the two major trajectories of enthusiasm on the one side, and skepticism and rejection on the other side.

A first strand of the literature[45] has strongly welcomed the integration of technology in corporate structures and has considered technologies as a panacea of the "evil" corporations' internal functioning, first of all conflicts of (human) interests. It has been said that the new digital means can reduce the risks of conflicts of interest in corporate relationships, in particular, between shareholders and directors, while enabling a more direct interaction between the two different parties to the social contract, and thus greater control by the former onto the latter. In accordance to these views, the reduction of agency problems, would nullify, at their very roots, some of the corporate tools that have been introduced as a remedy, exactly for the

---

[41] R. WILE, *A Venture Capital Firm Just Named An Algorithm To Its Board Of Directors — Here's What It Actually Does*, *Business Insider*, 2014, https://www.businessinsider.com/vital-named-to-board-2014-5?r=US&IR=T#:~:text=Just%20like%20other%20members%20of,a%20specific%20company%20or%20not.

[42] N. ABRIANI-G. SCHNEIDER, *Diritto delle imprese e intelligenza artificiale*, Il Mulino, 2021, 197 ff.

[43] G. MOSCO, *AI and the Board within the Italian Corporate Law: Preliminary Note*, in European Company Law Journal, 2020, 3, 87 ff., 89 and 91-92.

[44] O. BORGOGNO, *Making decentralized autonomous organizations (DAOs) fit for legal life: mind the gap* (October 2022), Questioni di Economia e Finanza, Banca d'Italia https://www.bancaditalia.it/pubblicazioni/qef/2022-0718/QEF_718.pdf., especially 7-14.

[45] See lately, Z. LI, *ArtificialFiduciaries*, U. of Pittsburgh Legal Studies Research Paper No. 2023-31*, 18-28.* D. YERMACK, *Corporate Governance and Blockchains*, 21 REV. FIN. 1, 9 (2017), 17-20.

purposes of mitigating conflicts of interests between shareholders and directors, or between different categories of shareholders, as independent directors.

Following these lines of reasoning, technologies, namely both blockchain and artificial intelligence systems, are said to have the capabilities of radically changing the structure of corporations as we know them today, because these tools have the potential to more efficiently link the different groups of interests orbiting around a corporation, minimizing the risks of self-dealing behaviours, and are thus apt to reach the optimal allocation of financial resources within the new corporate "machine"[46]. In line with these views, the advancements brought about by new technologies should be promoted just for the purposes of revising those behavioural externalities of the traditional hierarchical corporate structures: technologies open up indeed the ways to horizontal (or platform-based) governance models which are deemed to be more efficient than the so-far known corporate governance solutions.

This first approach of "corptech-enthusiasts" supports radical evolutions of traditional corporate structures into forms that either strongly depart from the corporate anatomy, as occurs with decentralised autonomous organizations[47], or that proposes a maximum integration of technology in the corporate dynamics, up to the point of admitting the presence of robo-directors, that is artificial intelligence systems appointed as directors[48].

---

[46] I. EREL-L H. STERN- C.TAN- M.S. WEISBACH, *Selecting Directors Using Machine Learning*, European Corporate Governance Institute (ECGI) - Finance Working Paper No. 605/2019 (2018), 3226 ff., 3227; A. HAMDANI- N. HASHAI- E. KANDEL-Y. YAFEH, *Technological Progress and the Future of the Corporation*, 6 J. BRITISH ACAD. (2018) 215 ff., 223-233.

[47] DAOs are not corporations, DAOs are new organizational models for conducting business that are different from corporations. E. DI MARTINO-O. BORGOGNO, *Decentralised Autonomous Organizations: Targeting the Potential Beyond the Hype*, EBI Working Paper 2024-161, especially at 24-32.

[48] This has happened with the case of the algorithm "Vital". This solution would not be possible in the European Union, where AI systems lack legal personality. See G. TEUBNER, *Digital Personhood? The Status of Autonomous Software Agents in Private Law* (11 May 2018) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3177096, 5-7.

Corptech proponents are accused by a second strand of the literature of being blended by a "tech nirvana fallacy"[49]. The fallacy in which these commentators incur is related to the fact that technologies are merely new instruments through which traditional corporate relationships may unfold: in accordance with this conception, technology is not going to change anything of the corporate structure, not even the common conflict of interests situations among their parties. To the very contrary traditional agency problems are very likely to be propagated, and in some cases even worsened through algorithmic codes.

This second position of technology "scepticists" may turn to solutions of rejection of technology in the internal corporate governance sphere, in light of the various technology-related risks that are growing together with the fast technological development and that are often not fully tackled by the slower evolving regulatory framework. The consideration of the open-ended nature of these risks, may suggest that the costs of technological transformation of corporations is higher than its expected benefits, thus discouraging corporations to take this path.

As summarised, the current debate on corptech appears to be polarized around two different extremes: extreme welcoming or extreme suspicion regarding the digital evolution of corporate governance. At a closer consideration, the current debate analyses the possibilities of integrating technologies in the corporate governance domain, in light of the traditional corporate governance problems (agency problems) and tools (the tools restoring these same problems).

Although insightful, the level of the debate developed so far appears to remain confined in too general terms: it indeed considers AI as an alter-ego of human being, this further fueling a subsequent polarization: to put it in

---

[49] L. ENRIQUES — D.A. ZETZSCHE, *Corporate technologies and the tech nirvana fallacy*, in *Hastings LJ* 2020, 72, 55 ff., 71-74.

extremely radical terms, either AI is to be tendentially fully admitted as a player into corporate structures (100% relevance) or, although it may bring about efficiencies in corporate. decisions, it won't overall change anything in respect to traditional corporate governance mechanisms (0% relevance). In the end, this framing of the relationship between AI and corporate structures, does not seem to provide concrete answers regarding which should be the degree of corporations' internal digital transformation admitted under the present IT and corporate governance framework.

From this more specific standpoint, the major flaws of the present corptech debate relate to the fact that such debate assumes corptech as a unique technological product and does not delve into the heterogeneity of corptech tools, not only given by the variety of possibly applicable technologies, but – what is relevant for the purposes of our study – as given by the different roles corptech tools can have at the various levels of the corporate governance structure.

In line with the conception of corporate governance as an organizational underpinning for a more efficient decision-making processes in corporate matters[50], these different levels are to be first of all identified on the basis of the different types of decisions that the board of directors takes in a corporation. Hence, the role of corptech tools must be measured with the *type* of corporate decisions to which these are employed for.

4. *Corporate Decisions in Between Procedures and Discretion*

In our view, the above pictured scholarly debate misses to assign a specific– and more realistic – role to AI (in between the two extreme hypothethical options of 100% and 0%) because it misses to consider the diversity of corporate decisions, in which AI can be employed.

---

[50] J. ARMOUR ET AL., *The Anatomy of Corporate Law: A Comparative and Functional Approach*, 3ª, Oxford, Oxford University Press, 2017, 22-24.

We propose a different framework, starting from the variable of completeness and discretion involved in a given decision-making process. For the sake of clarity, we can take two opposite cases.

A first hypothesis might relate to a constrained decision, as it occurs in the case of the already mentioned high frequency trading system. In this case, the machine applies a protocol (that is, a set of criteria ex ante decided), which provides a precise formalization of the decision-making process and which is thus validated by established rules. Exactly because a protocol exists, these decisions are thus bounded in their patterns and outcomes, because the protocols are normally based upon the same (quantitative) variables.

This is why it can be argued that decisions regarding businesses organizational structures are performed in a "complete" environment: indeed, the elements upon which the decision has to based are known and only have to be assembled so as to identify the decision's outcome. For these decisions, automated systems appear to have a better performance in the moment of the execution of the protocol. Minor discretion is here left to the machine, because a (relatively low) discretion has been already excercised by the human in the moment of the definition of the protocol/procedure. If the score is above a certain threshold, then an action is performed. These decisions can be recurring and continuous: they are indeed about the processing of multiple data in accordance with a pre-determined protocol, which is sufficient to fully substantiate the decision-making outcome. This is what occurs, for example in the case of high frequency trading, of the processing of information for AML enquiries or of the distribution and surveillance of employees.

These examples relate to limited-scope decisions, that we may call "organizational routine" or "complete" decisions", elsewhere defined as

high frequency low impact decisions[51]. Involving low discretion because the outcome is binded to a given protocol. In these cases, the AI device substitutes the human decision-maker only at the secondary stage of executing a protocol; a task that the device can do faster, at a cheaper cost and with greater matemathical precision.

This type of recurring decisions can be taken in nearly fully automated way[52], provided the human factors is capable of intervening as a supervisor or controller of the enacted system, majorly in terms of monitoring and reviewing of the same system's output, just as the recalled normative requirements demand.

In these cases, accountability of the decision-making processes can– or better said should– be reached through the product-based approach underlying the "human in the loop" commands: the decision loop is here majorly given by the machines' output, and the role of the human decision-maker is to ascertain whether the machine-driven outputs are suitable or fair (e.g. free of biases) in respect to the framing ex ante developed in the protocol. Ultimately, in respect to routine and complete decisions, the human-in-the-loop paradigm may be sufficient to guarantee legitimacy of the decision: having the protocol, it is easy to check whether the machine has applied it correctly. The human subject is thus capable of exercising oversight functions of the model, to which the decision-making loop is left on efficiency grounds.

From the perspective of corporate structures, hence, it can be said that in respect to routine decisions, AI directors might employ fully automated

---

[51] A. CAMUFFO — A. GAMBARDELLA — A. PIGNATARO, *Framing Strategic Decisions in the Digital World*, in *Strategic Management Review* 2023, 4, 2, 127 ff., 131-133. See also M. PETRIN, *Corporate Management in the Age of AI*, in *Columbia Business Law Review* 2019, 2019, 3, 966 ff., 983. See also V. KOLBJØRNSRUD — R. AMICO — R. THOMAS, *The promise of artificial intelligence: Redefining management in the workforce of the future*, Accenture Institute for High Performance, 2016

[52] A. CAMUFFO — A. GAMBARDELLA — A. PIGNATARO, *Framing Strategic Decisions in the Digital World*, in *Strategic Management Review* 2023, 4, 2, 127 ff., 131-133.

decision-making processes and thus rely on fully automated outputs, provided that corporate structures enable, in addition to an efficient automated process, also a risk-based monitoring function. In this respect, it must be noted that the oversight function is not always effective in respect to the risk-minimization objectives in the context of organizational decisions: as some recent empirical studies have demonstrated, indeed, the human monitoring function is better suited to the detection of smaller machine-driven mistakes, than for the fixation of larger and more severe ones, this rendering the human oversight function not always "secure" also in respect to highly repetitive, thus predictable decisions[53].

5. *The Case of Strategic Decisions in Incomplete Scenarios*

At the other end of the spectrum we find strategic decisions, as those involving corporations' fundamental changes (e.g. a decision regarding a merger with another company). This is a decision taken in an incomplete environment: the protocol is not precise by hypothesis.

From the decision-maker's standpoint, this type of decision involves a number of potentially infinite framing options, which respectively might lead to different outcomes[54]. Incompleteness makes the formalization of the decision at stake difficult, and with that the construction of a one-sided protocol[55].

Strategic decisions can indeed be framed differently, and as a result may lead to different outcomes. It all depends on the functional standpoint and thus the sensitivity with which these are taken. In these decisions, indeed,

---

[53] D. SELE — M. CHUGUNOVA, *Putting a Human in the Loop: Increasing Uptake, but Decreasing Accuracy of Automated Decision-Making*, (SSRN Scholarly Paper), Rochester, NY, 2022, 14.

[54] S.A. ALVAREZ — J. PORAC, *Imagination, Indeterminacy, and Managerial Choice at the Limit of Knowledge*, in *Academy of Management Review* 2020, 45, 4, 735 ff.

[55] T. ERIKSON- M. KNOCKAERT, *Negotiating Incomplete Contracts*, in *Journal of Strategic Contracting and Negotiation*, 2021, 5, 3, 153 ff..

there's no right mathematical answer, but a decision output will depend on how different qualitative factors are evaluated and "assembled".

Against this backdrop, for the purposes of the present analysis, a strategic decision assumes incompleteness, and incompleteness grounds the discretionary nature of a decision-making process.

In the field of corporate action incompleteness appears to be first of all directly related to the complexity of the interests at stake in corporate governance: the so-called compositive nature of corporate governance, is becoming increasingly relevant in times where corporate efficiency goals are object of a heated debate and of reform projects requiring boards to take into consideration different and "diverse" elements for the purposes of their decision-making.

In second stance, incompleteness of the setting of strategic corporate decision-making is given by the unpredictability and uncertainty of business relations, subject to sudden health or geopolitical crises, which require an additional effort of flexibility of corporate governance.

As given by the considered determinants, strategic decisions structurally involve an area of discretion. The degree of discretion involved in a decision can evidently be varied, along the lines of a continuum in which low discretion decisions are more similar to routine tasks and high discretion decisions correspond to truly strategic decisions. The greater the degree of discretion is, the more difficult it becomes to properly allocate decision rights between the human decision-maker and the machine. In the absence of a protocol, indeed, the machine evidently cannot perform the execution task mentioned in the previous paragraph.

Nonetheless, also in these decisions, AI tools may facilitate the attainment of an efficient decision-making process: this can occur through the delegation to the machine of specific sub-tasks of the overall decision-making process.

Ideally algorithms can also perform the task of framing a protocol for strategic decisions, as is the case of an M&A plan defined through automated means.

The developments in the field of machine learning clearly show the creative potential of last generation AI models in creating a decision-making process from the start of the framing of a given problem. The range of problems addressed by machine learning typically involves tasks for which other approaches fail, either because there is no suitable formalization of the problem, or because the resolution of the problem is intractable with non-learning approaches.

Indeed, machine learning typically refers to algorithms that learn to complete tasks by identifying statistical patterns in data, rather than following instructions provided by humans. Machine learning approaches focus on the development of systems capable of learning and inferring from data to solve an application problem without being explicitly programmed with a set of step-by-step instructions from input to output. Learning refers to the computational process of optimizing from data the parameters of the model, which is a mathematical construct generating an output based on input data[32].

However, although feasible from a technical standpoint, the delegation to an AI system of a strategic decision needs to be better assessed from a legal standpoint, in light of the very features of such a decision, namely incompleteness and discretion. The challenge is to justify the employment of AI in strategic decisions not only upon an efficiency rationale but also identifying its legitimacy grounds in the specific context of corporate decisions, which are to be considered in addition to and provided that basic human monitoring requirements are fulfilled.

From this different perspective, the factual feature of incompleteness appears to challenge the effectiveness of the current regulatory approach

based on a human monitoring function, as the one outlined by the AI Act, when it comes to strategic corporate decisions.

The role of the human decision-maker as a mere monitoring principal of the machine-agent leaves appears to be highly risky in the context of strategic decisions. Exactly because these decisions come to operate in incomplete environments, they may be conducted without any direct prejudice to a legally relevant interest, as a fundamental right, but they could still result to be erroneous in the sense of not indicating the "better" decisional outcome for the corporation that addresses it.

A protocol may be needed to execute a similar decision, but will never be sufficient to govern all the (potentially infinite variables) of a strategic decision. As a result, also the oversight over whether the protocol has been correctly implemented by the machine is not sufficient of "monitoring" a strategic decision.

Last but not least, the monitoring paradigm informing current IT-related regulations is not capable of ensuring accountability of human actors' strategic decisions conducted with the support of AI. In these cases, there is indeed not only the problem of AI's monitoring (a problem with which the law struggles), but also of the accountability of how AI is used in the decision-making process (a problem that the law has not yet tackled).

Under these premises, the following paragraphs will enquire the *internal* limits which corporate governance poses to the integration of its traditional structures– and in particular the management– with digital components with specific regard to strategic decisions. Consideration will be ultimately given to the impact on directors' duties when AI tools are employed for the purposes of strategic decision-making.

Hence taking into consideration a corporate strategic decision, characterized by its high impact and its large degree of incompleteness (as the appointment of a board member or an M&A prospectus), the following

analysis will be divided into two major parts: at first, consideration will be given to a corporate strategic decision solely based on AI and its decrease in accuracy; in second stance, the human monitoring capabilities in respect to this type of decision will be questioned.

SECTION III – Strategic Decisions Solely Based on AI

6. *Strategic Decisions in Incomplete Environments and AI: Losing Accuracy*

As previously mentioned, integrating AI into decision-making processes is primarily motivated by efficiency considerations[56].

In routine decision-making, an automated system is expected to execute a task and provide an exact outcome based on a predefined protocol. The exact outcome is the crucial goal: in a limited discretion decision, its pursuit can be effectively entrusted to the algorithm, with the human being placed to guard (not against the exact outcome supposedly granted by the algorithm, but) against any negative externalities impacting the (fundamental) rights of individuals[57].

Therefore, in routine decisions, the accuracy of the AI tool seems an implicit but powerful source of its legitimacy. The algorithm serves as a self-sufficient and potentially exclusive decision-making strategy for corporate routine decisions. In accordance with the HITL paradigm, human involvement is limited to overseeing the AI system in use. Intervention is

---

[56] G. MALGIERI — F.A. PASQUALE, *From transparency to justification: Toward ex ante accountability for AI*, in *Brooklyn Law School, Legal Studies Paper* 2022, 712, 12; J.A. KROLL ET AL., *Accountable Algorithms*, in *University of Pennsylvania Law Review* 2017, 165, 3, 636, who states "the efficiency and accuracy of automated decisionmaking ensures that its domain will continue to expand".

[57] Consider the example of an algorithmic selection that is correct in itself but discriminatory with respect to gender, such as the greater reliability of the female gender in loan repayment. The result is correct, but inadmissible due to the principle of non-discrimination. See H. MATSUMI — D.J. SOLOVE, *The Prediction Society: Algorithms and the Problems of Forecasting the Future*, (SSRN Scholarly Paper), Rochester, NY, 2023, 1 ff.

necessary only in emergencies to protect fundamental rights as defined by applicable regulations.

The implicit assumption behind these assertions is that humans might refrain from intervening in the algorithmic decision-making process because, as long as the results are highly reliable, the high costs associated with ongoing human intervention may not be justified. If the process produces accurate results, there may be no need to structure monitoring instruments, relying instead on remedies designed to preemptively address any negative externalities related to health, safety, human rights, and the environment rather than scrutinising the decision-making process. In essence, the machine is entrusted with finding the optimised and most efficient solution while humans manage any side effects impacting critical areas.

However, this equilibrium is no longer applicable when dealing with strategic decisions.

In such cases, incompleteness implies that the exact outcome of the decision is not entirely within the control of the decision-maker, whether human or algorithmic. Unlike routine tasks, strategic decisions inherently entail a margin for error due to the incomplete nature of available information. Even the most high-performing AI cannot comprehend everything, and the broader the incompleteness, the less accurate a decision based solely on AI becomes.

The issue extends beyond less accurate calculations; as noted, incompleteness not only results in an inevitable reduction in accuracy but also introduces the potential misplacement of the core idea on which the use of the algorithm is based – identifying the right decision with the accurately optimised decision (§ 3). Incompleteness may signify not just a scarcity of observations within a given variable (such as the number of sick leave days for director X in the next three years) but also, and more

fundamentally, the inadequate consideration of certain variables (e.g. the algorithm fails to properly weigh the ability to work together between director X and director Y, thereby neglecting or significantly underestimating that variable)[58]. In the latter case, incompleteness means that an algorithmic optimisation, even if computationally exact, would not necessarily lead to the "right" outcome[59].

Therefore in decisions taken in structurally incomplete settings, the genetically diminished statistical accuracy is compounded by the compression of the margins of "legal" accuracy, which stems from an organised hierarchy and systematisation of legally relevant interests and regulatory objectives insisting on a given decision-making framework. Consequently, the more a decision depends on an incomplete factual matrix, the more challenging it becomes to reconcile individual and collective interests with the "system" (specifically, an algorithmic system) that is impossible to identify ex-ante in a protocol.

Hence, a fundamental tension arises between statistical prediction and "legal" accuracy in scenarios with incomplete risks, as is often the case in business activities.

This inevitable expansion of uncertainty renders the algorithmic result doubly unreliable, and effectively managing the incompleteness associated with algorithmic decision-making in a strategic context becomes crucial.

7. *"A Soul to Damn and a Body to Kick": AI Decisions and the Legal Strategies*

---

[58] D.J. SOLOVE — H. MATSUMI, *AI, Algorithms, and Awful Humans*, (SSRN Scholarly Paper), Rochester, NY, 2023, 7.

[59] According to SOLOVE — MATSUMI, *AI, Algorithms, and Awful Humans* (n. 58) 7, "problems emerge when too much quantitative data is relied upon to the exclusion of qualitative data. Not everything is readily quantifiable".

With its innate confrontation with an incomplete reality, the law has developed multiple tools and strategies to deal with this issue.

The challenges associated with managing strategic decision-making are particularly acute in the context of corporate regulation. The complexities of overseeing organisations entrusted with long-term productive activities necessitate corporate law to address the management of strategic decisions due to their inherent uncertainty[60].

A widely held and effective interpretation of corporate law posits that it has two primary objectives: first, to establish the structure of the corporate form, and second, to minimise or eliminate conflicts of interest among various corporate stakeholders[61].

In an ideal world of perfect decision-making, market logic might achieve the second objective through an efficient negotiation between principals and agents[62]. However, in the real world, subject to market failures, developed corporate systems employ sophisticated strategies to allocate decision-making discretion. Dealing with future events, most of these legal strategies refrain from pursuing the exact outcome of a certain decision. Instead, they focus mainly on the decision-making process as a key area to mitigate agency problems, enhancing the principal's ability to control or structure the agent's decisions through (agent constraint, incentive alignment, decision rights, etc.)[63].

---

[60] See in this regard the analysis by L. ENRIQUES, *Pandemic-Resistant Corporate Law: How to Help Companies Cope with Existential Threats and Extreme Uncertainty During the Covid-19 Crisis*, ECGI Law Working Paper 530/2020.

[61] See *ex multis* J. ARMOUR — H. HANSMANN — R.H. KRAAKMAN, *Agency Problems and Legal Strategies*, in *The Anatomy of Corporate Law: A Comparative and Functional Approach*, , 3ª, Oxford, Oxford University Press, 2017, 29 ff.

[62] On the difference between efficient and desirable allocation of resources, see J. ARMOUR ET AL., *Principles of Financial Regulation*, Oxford, New York, Oxford University Press, 2016, 54 ff.

[63] P. DAVIES, *Introduction to Company Law* (Book, Whole), Oxford, Oxford University Press, 2020, 30.

Given these legal strategies and postulating their effectiveness, the question then arises whether they can be applied to decisions based solely on AI to compensate for the mentioned loss of algorithmic accuracy.

An interesting vantage point from which to investigate this issue is provided by the Vital case, where – according to media reports – an algorithm was granted voting rights on the board of a Hong Kong venture capital firm, when it comes to deciding "whether the firm makes an investment in a specific company or not"[64]. Legally speaking, Vital was not technically a corporate director under Hong Kong's corporate law. It was, however, considered "as a member of [the] board with observer status" by the other human board members and in that capacity "contributed" to approve some investment decisions[65].

Despite the AI's involvement in decision-making, the complex duties and responsibilities associated with the office remained with the human directors[66]. The AI functioned as a "narrow-gauge" director, while the human directors played a role akin to caretakers for the digital prodigy, being liable for its limits and shortfalls.

---

[64] WILE, *A Venture Capital Firm Just Named An Algorithm To Its Board Of Directors – Here's What It Actually Does* (n. 41).

[65] N. BURRIDGE, *Artificial Intelligence Gets a Seat in the Boardroom*, *Nikkei Asia*, 2014. In the legal literature on this topic see *ex multis*, F. MÖSLEIN, *Robots in the Boardroom: Artificial Intelligence and Corporate Law*, (SSRN Scholarly Paper), Rochester, NY, 2017 1; S.A. GRAMITTO RICCI, *Artificial agents in corporate boardrooms*, in *Cornell L. Rev.* 2019, 105 871 s.; G.D. MOSCO, *AI and the Board Within Italian Corporate Law: Preliminary Notes*, in *European Company Law* 2020, 17, 3, 89; M. EROĞLU – M. KARATEPE KAYA, *Impact of Artificial Intelligence on Corporate Board Diversity Policies and Regulations*, in *European Business Organization Law Review* 2022, 23, 3, 541 ff.

[66] GOYAL, *Hong Kong VC firm appoints AI to Board of Directors,* in *IT Business Blog*, 16 maggio 2014, in https://www.itbusiness.ca/blog/hong-kong-vc-firm-appoints-ai-to-board-of-directors/48815 (Consultato: 29 gennaio 2024]. On this topic, see also MOSCO, *AI and the Board Within Italian Corporate Law* (n. 65), 92, who noted that "not only is it currently undesirable for AI-based directors to sit in on the board, but times are also unripe for even the most preliminary discussion as to AI-related subjectivity and legal capacity. This, in turn, makes it impossible to conceive of AI as a truly independent entity – one that could be distinguished from human directors".

In the perspective considered here of entrusting a strategic decision to AI, this example highlights two critical issues: AI is indeed a) unable to be the legal recipient of either civil or criminal liability and, for what is particularly relevant in this context, b) impervious to the set of duties typical of a director of a corporation.

That a machine cannot be held legally accountable or subject to the same set of duties as a corporate director is a well-known issue with no easy solution. Simply changing the law to allow for the appointment of a self-learning algorithm as a director would not be enough, as the role of a director is rooted in the legal concept of accountability[67].

In this sense, starting from the practical observation that AI lacks "a soul to damn and a body to kick"[68], it has been correctly observed that "governance relies on accountability, accountability presupposes a conscience, a conscience might presuppose consciousness"[69]. Although it cannot be ruled out that one day AI may become advanced enough to develop or at least imitate such human characteristics, this is not yet the case[70].

Since "accountability requires more than legal capacity"[71], simply allowing today's AI into the boardroom would ultimately produce a shift of liability onto humans, as seems to have happened in the case of Vital[72].

However, it is unnecessary to go that far (i.e., search for the consciousness of AI) to highlight the tensions between current corporate regulation and AI decisions.

---

[67] See GRAMITTO RICCI, *Artificial agents* (n. 65), 886.

[68] The original quotation from Lord Thurlow was about corporations: "Corporations have neither bodies to be punished, nor souls to be condemned, they therefore do as they like".

[69] GRAMITTO RICCI, *Artificial agents* (n. 65), 894.

[70] GRAMITTO RICCI, *Artificial agents* (n. 65), 894 and 906, also noted that such a situation, once occurred, would likely pose a series of more existential problems than those – albeit important – encountered in regulating a board.

[71] GRAMITTO RICCI, *Artificial agents* (n. 65), 906.

[72] Correctly GRAMITTO RICCI, *Artificial agents* (n. 65), 906, noted that "proposals that emphasized the role of insurance in order to repair damages caused by artificial agents in boardrooms exclusively consider ex-post remedies that aim to repair already caused damages. Such proposals would fail to address or enhance accountability itself".

The passage regarding the inability of an algorithm to fulfil all the duties imposed by the legal system on every director is also, if not more, significant.

The issue concerns not only the constraints imposed on directors but, more broadly, many of the legal strategies used by the most sophisticated legal systems to manage agency problems and, more generally, market failures. When the point of application of current corporate legal strategies shifts from the human decision-maker to the algorithmic one, the effectiveness of these strategies cannot be taken for granted.

The main corporate law strategies are structurally based on the specific characteristics of human beings, which cannot consistently be replicated or even exhausted in an automatic optimisation process. This does not mean resorting to the irrational or instinctual side of humans, which is unknown to machines: in order to do so, it would first be necessary to demonstrate that human irrationality solves more problems than it creates[73]. Rather, the current corporate law system has built legal strategies aimed at exploiting dynamics that are (not good or bad in themselves, but) quintessentially human. In this sense, these strategies are designed as structurally immanent and functional to the human nature of their recipients. Therefore, their immediate applicability to algorithms appears to be anything but obvious, given the radical difference in decision-making processes between algorithms and humans[74].

---

[73] After the global endorsement of behavioural economics and its conclusions, it seems increasingly difficult to imagine that we can rely directly on our own biases. See *ex multis* T. GILOVICH — D. GRIFFIN — D. KAHNEMAN, *Heuristics and biases: The psychology of intuitive judgment*, London, Cambridge university press, 2002; D. KAHNEMAN, *Thinking, fast and slow*, New York, Farrar, Straus and Giroux, 2011; C.R. SUNSTEIN, *Empirically informed regulation*, in *U. Chi. L. Rev.* 2011, 78, 1349 ff.

[74] SOLOVE — MATSUMI, *AI, Algorithms, and Awful Humans* (n. 58) 7, according to whom "machine decisions are fundamentally different from human ones. Comparing human to machine decision-making is akin to comparing apples and oranges, not rotten apples to fresh ones".

Two examples can be helpful to highlight this difference concerning the board of directors.

7.1. *Human Versatility and the Limited Operational Scope of AI*

Human directors are inherently versatile subjects, capable of interacting and reasoning about problems with different characteristics. Therefore, they can apply their sensitivity to a wide range of decision-making processes. This versatility has recently been seen as a distinctive and positive element within the boardroom, particularly against a trend towards the hyper-specialization of human directors[75].

On the contrary, both well-known operational examples, such as the aforementioned Vital, the DAO experiment and algorithms used in mergers and acquisitions[76], as well as legal scholars converge on the narrow "operational domain" of the algorithm[77]. It is worth noting that this assertion is not contradicted today by LLMs (Large Language Models), whose progress in the language domain has attracted public attention in the past year. Still, LLM content is not as reliable as human-generated ones[78].

Even when considering cases of autonomous governance intelligence, the current conclusion seems to be that "the goals of the AI system and the

---

[75] See R. SHAPIRA — Y. NILI, *Specialist Directors*, (SSRN Scholarly Paper), Rochester, NY, 2023.

[76] Having proposed at least six type of different functions performed by algorithms, M. GAL — D.L. RUBINFELD, *Algorithms, AI and Mergers*, in *Antitrust Law Journal* 2023, forthcoming, noted that "this range of uses also exemplifies the fact that there is no one-type-fits-all algorithm. Rather, different types of algorithms are needed to perform different tasks".

[77] F. MERTENS, *The Use of Artificial Intelligence in Corporate Decision-Making at Board Level: A Preliminary Legal Analysis*, (SSRN Scholarly Paper), Rochester, NY, 2023, 12.

[78] At present, the testing of LLM in total autonomy in the face of complex problems has made headlines more for extravagant results than for robust returns. However, it must be emphasised that the ability to handle language, the medium through which humans share their logic, appears to be one of the most interesting frontiers in the development of AI towards greater versatility.

company are restricted to a narrow operational domain, in addition to being closely linked and intertwined"[79].

The underlying implications are significant. It has been correctly observed in legal contexts that discussions about algorithmic decision-making often imply a division of labour between AI and humans, based on an ideal division of tasks according to their respective competencies (machines are better at this / humans are better at that)[80].

If not properly managed, this division risks losing sight of the complexity of actual cases, focusing only on the advantageous aspect of the algorithm and leaving all other aspects to an overwhelmed human director. In essence, it echoes the Vital case, highlighting the gap between the machine's narrow domain of operation and human versatility.

7.2 *Collegiality and Diversity*

Corporate governance utilises the versatility of human directors (also) for monitoring, balancing, and transparency, for example, by requiring certain decisions to be made not by the individual CEO but by the board as a whole. This introduces a significant difference compared to algorithmic decision-making, where the algorithm optimises, but the board weighs. The decision to shift responsibility to the board level allows for solutions that result from reconciling different positions and weighing diverse perspectives. Replicating this dynamic is complex, not only when entrusting the decision to a single AI but also when imagining a plurality of algorithms interacting with each other. Indeed, a collegial decision is not simply the average of

---

[79] MERTENS, *The Use of Artificial Intelligence in Corporate Decision-Making at Board Level* (n. 77), 13.
[80] CROOTOF — KAMINSKI — PRICE II, *Humans in the Loop* (n. 19), 460 ff.

each board member's optimisations but rather a creative process through which the decision takes shape[81].

It has been observed that simply replicating the current strategy for human gender diversity through an algorithm without requiring actual representation of both genders on the board would be incorrect[82]. This approach would fail to acknowledge that diverse and original perspectives are shaped by the same incentives and disincentives, which are inherently different in the case of algorithms[83]. Algorithmic logic operates within a framework of paradigms that are not directly comparable to human ones and is not motivated by the same incentives that affect human decision-making. Equating algorithmic opinions with those of a human director is an erroneous approach due to the fundamental differences between their decision-making processes.

8. *The Contraction of Space for Strategies and the Loss of Control over the Decision-making Process*

The core of the argument lies in the explicit design of current corporate law strategies for a human decision-making process aimed at mitigating risks arising from agency relationships and the consequent threat of opportunism on the agent side. CorpTech itself was initially proposed as a potential tool to alleviate agency problems[84].

For the reasons highlighted above, stemming from the absence of a "soul to damn and a body to kick" to the "foreign" logic of AI, existing corporate

---

[81] KOLBJØRNSRUD — AMICO — THOMAS, *The promise of artificial intelligence* (n. 36), 13.

[82] On this topic EROĞLU — KARATEPE KAYA, *Impact of Artificial Intelligence on Corporate Board Diversity Policies and Regulations* (n. 65), 541 ff., 565.

[83] But *contra* see PETRIN, *Corporate Management in the Age of AI* (n. 36), 1003, according to whom "the combined knowledge and skills, benefits of group-decision making, and characteristics
such as diversity and independence, which previously could only be offered by a collective, will be replicated in fused
boards through an algorithm's coding features".

[84] See supra note 44 and 45.

law strategies prove to be less effective in addressing challenges that may arise from the use of AI for strategicdecisions.

Consequently, reintroducing these strategies in a context where discretion is assigned to the algorithm appears to miss the final goal, i.e. the control over the discretion of the AI's decision-making process. As a matter of fact, for each decision-making process, corporate law typically provides a certain number of strategies, usually proportionate to the significance of the decision-making process in terms of its impact on society, agency risks, effects on stakeholders, etc. However, to implement such legal strategies, there is a need for legal "space". An eloquent example might be the separation of powers, a foundation of every liberal democracy[85]. More pragmatically, within the corporate realm, one could highlight the difference between a decision made by the board of directors and one made by a sole director.

While, in the former case, activities like gatekeeping can be implemented using legal strategies such as trusteeship or diversity within the board or through a reconstruction of how the decision was reached via an analytical record of the board meeting, in the latter case, the legal "space" for the preparation of these strategies is unavailable as the entire process remains within the mind of the sole director, residing in a region impervious to legal strategies.

Similarly, if the algorithm proves to be as impervious to legal strategies designed for human directors, the consequence is that the fraction of the decision-making process entrusted to the algorithm lacks safeguards, reducing the overall legal space available to mitigate potential issues.

In summary, the case of strategic corporate decisions taken in incomplete scenarios highlights how the ineffectiveness of many current legal strategies

---

[85] The modern theory of distribution of powers is commonly ascribed to C. DE S. BARON DE MONTESQUIEU, *De l'esprit des loix*, vol. 1, Aux dépens de la Compagnie, 1749.

compounds the inevitable reduction in the accuracy of the algorithm. The outcome is a loss of reliability in the algorithmic result, unbalanced by remedies effectively addressing the process leading to that result.

To draw a comparison, the situation is not markedly different from the scenario in which, in a self-driving car, the algorithmic guidance system disengages just before impact with an obstacle, leaving the human driver to manage a situation irreversibly compromised[86]. At that point, even if the driver behind the wheel were the best possible pilot (i.e., a plastic embodiment of all human strategies), the outcome could not change: the available space would be insufficient.

9. *The Human in the Loop Overloaded*

Drawing on the example of autonomous driving cars, it might be argued that guiding a vehicle, which involves a continuous flow of actions and real-time corrections, fundamentally differs from making decisions relevant to the community. In such decisions, humans are expected to have ample time to make the ultimate decision. However, unfortunately, when it comes to human-machine interaction, reality seems to offer less optimistic insights.

This is evidenced by the "human in the loop" approach, conceived as a human safety valve to "prevent or minimise risks to health, safety, fundamental rights, or the environment that may arise when a high-risk AI system is used" (Article 14 AI Act). Despite being the preferred strategy for decisions solely based on AI, the text proposed by the regulation has raised concerns, even without considering strategicdecisions.

Criticism has been directed at the inadequate attention provided by Article 14 to the human user's role, skills, and duties. The draft AI Act's human

---

[86] The case is far from theorical: see CROOTOF — KAMINSKI — PRICE II, *Humans in the Loop* (n. 19), 438.

oversight design requirements for providers are noted to "focus less on understanding and mitigating known human frailties or designing an effective human-machine system than on increasing the agency and power of the human in the loop"[87]. The absence of clear procedural safeguards related to the specific monitoring functions assigned to the human controller opens the door to a significant risk of overloading the human controller. Such potential overloading, from a factual standpoint, diminishes monitoring effectiveness and, from a legal perspective, makes it exceedingly challenging to verify when and how such functions have been diligently performed. Consequently, the supervising natural person would be endowed with broad theoretical powers, and subject to duties beyond her actual capabilities.

The situation becomes even more complicated in the context of strategic decision-making. In the realm of these decisions, human oversight, as described in Section 14 of the AI Act, appears insufficient to address the increased margin of error.

While the outlined regulatory developments provide a crucial regulatory baseline for delineating corporations' efficiency choices in business and digitalisation transformation matters, limited analysis exists on how the consolidating' human in the loop' paradigm tackles the risks that arise from the structural incompleteness features of AI-driven decision-making in a strategic context, beyond oversight tasks focused on fundamental rights protection.

In this scenario, merely adding inaccuracy to the list of risks to be monitored according to Art. 14 AI Act does not resolve the issue. Although monitoring can help mitigate negative externalities, achieving complete control over the potentially infinite variables involved in discretionary decision-making is a much more complex task. In other words, the issues raised on Art. 14

---

[87] CROOTOF — KAMINSKI — PRICE II, *Humans in the Loop* (n. 19), 504.

AI Act concerning human oversight seem destined to worsen when monitoring shifts from mitigating negative externalities related to fundamental rights to reviewing the intrinsic correctness of the whole decision-making process. The human compensating for the loss of reliability in the algorithmic outcome results in an overextension of human oversight and, inevitably, an (even more severe) overload of the strategy, transforming it into something more akin to strict liability[88].

By saying that, we do not imply that the human-in-the-loop approach is inherently ineffective, but solely that, by its nature, its scope cannot be excessively extended, as would happen by imposing oversight on the intrinsic correctness of the algorithmic process.

In simpler terms, it is evident that the human-in-the-loop approach, as structured in Article 14 of the AI Act, does not serve as a valid safeguard against the loss of reliability in the algorithmic result in the case of strategicdecisions.

This assertion is further reinforced by automation bias in human-algorithm interactions. Automation bias refers to the human risk of letting the machine's dictates take over the decision-making process or, in other words, making the wrong decision based on what the machine has (wrongly) indicated as the correct outcome[89].

Automation bias has proven to be a pervasive phenomenon not necessarily mitigated by the dsignificance of the decision. Uncritical reliance on algorithmic decisions has occurred even in highly sensitive circumstances, such as criminal judgments. The reference is notable to the case of Compas,

---

[88] See SELE — CHUGUNOVA, *Putting a Human in the Loop* (n. 53), 16, who states, "in a more sophisticated system, the number of features incorporated into the automated recommendation may exceed human capacity. If this is the case, human monitors may have to rely on inferior (or at least limited) information when deciding on the adjustment. This imbalance may increase the risk of decreasing decision accuracy due to human intervention further".

[89] On empirical evidence of automation bias see SELE — CHUGUNOVA, *Putting a Human in the Loop* (n. 53), 4 ff.

decided by the Supreme Court of Wisconsin[90], where the algorithm's decision on the duration of the sentence in proportion to the risk of recidivism was accepted without real counterbalances or monitoring activities regarding the decision-making process. Implicitly, it formed the basis for a human decision justified ex-post by applying typical human decision-making rules. The result was an opaque algorithmic decision uncritically legitimised by the human decision-maker who accepted and justified the outcome ex-post[91].

Coordinating these considerations with Article 14 of the AI Act, it is observed that paragraph 4 of the provision highlights the risk of automation bias but does not address it. As correctly noted, such a choice shifts the management of the automation bias risk onto human oversight, becoming an additional element of overload for this strategy.

10. *Strategic Decisions and a Human-centric and Trustworthy AI*

The approach chosen by the AI Act for decisions solely based on AI becomes problematic as one moves from decisions dealing with low incompleteness to those dealing with increasing incompleteness. Specifically, within the realm of AI-human interaction, the HITL strategy can be considered a simplified approach applicable to particular cases. In these cases, the substantial human replacement resulting from decisions solely based on AI seems justified by the minimal or limited degree of discretion required for

---

[90] *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016). On the case, see *State v. Loomis*, in *Harvard Law Review* 2017, 103, 5, 1530.

[91] The ruling has been intensively criticised from different perspectives. Ex multis see J. MATTU ET AL., *How We Analyzed the COMPAS Recidivism Algorithm*, *ProPublica*, 2016; K. FREEMAN, *Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis*, in *North Carolina Journal of Law & Technology* 2016, 18, 5, 75 ff.; I.D.M. BERIAIN, *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling*, in *Law, Probability and Risk* 2018, 17, 1, 45 ff.; S.G. MAYSON, *Bias in, bias out*, in *YAle lJ* 2018, 128, 2218 ff.

the decision and the high level of accuracy achieved. In this sense, it would appear that current European regulations based on HITL envision a situation characterised by low discretion and a high reliability of the algorithmic outcome; this is likely because such a scenario has been prevalent in these early years of digital transition.

When delving into the domain of strategicdecisions, the regulatory architecture established by the AI Act for decisions solely based on AI appears inadequate to meet the aforementioned principal prerequisite of the European project, stating that "artificial intelligence should be a human-centric technology. It should not substitute human autonomy or assume the loss of individual freedom and should primarily serve the needs of society and the common good"[92].

The analysis here undertaken calls for a radical shift in the regulatory perspective. In the incomplete environment of strategic decisions, the normative goal cannot solely be an algorithm that does not endanger the safety of its users. Instead, the normative goal should be a human who can harness, while remaining in command, a tool that enhances their decision-making capabilities.

The final part of this work contemplates some guidelines for implementing what has been defined at the European level as the "human in command" approach within the realm of strategic decisions. According to this approach, "machines remain machines, and people retain control over these machines at all times"[93].

SECTION IV – AI as a tool

11. *A Ban of the Algorithm: A Wrong Turn*.

---

[92] AI Act, recital 4b.
[93] EUROPEAN ECONOMIC AND SOCIAL COMMITTEE, *Artificial Intelligence: Europe needs to take a human-in-command approach, https://www.eesc.europa.eu/en/*, 2017.

It is worth immediately clearing the field of what seems to be the simplest solution to restore the human in command, namely a preventive and absolute ban on using the algorithm for strategic decisions.

This approach can lead to undesirable outcomes. At the macroeconomic level, it is a losing proposition since it is evident that not embracing the race towards innovation today means suffering from the decisions of competing jurisdictions tomorrow[94].

Even without dwelling on the systemic opportunities that might be lost through refraining from embracing AI disruptive innovation, at the microeconomic level a ban on using AI risks turning out only a formal prohibition, impossible or extremely challenging to enforce in practice. A recent investigation has shown that AI's ability to save time and resources is significant enough to prompt a considerable percentage of generative AI users to use the tool at work, even without authorisation or against the employer's ban, while recognising that the safe and ethical use of this tool is through company-approved programs[95]. They may also pass off the machine's work as their own[96]. The probable outcome of such a ban on the

---

[94] The very text of AI Act stress both the importance of development of a human-centric and trustworthy AI. According to recital 3 AI Act, "artificial intelligence is a fast evolving family of technologies that contributes to a wide array of economic, environmental and societal benefits across the entire spectrum of industries and social activities. By improving prediction, optimising operations and resource allocation, and personalising digital solutions available for individuals and organisations, the use of artificial intelligence can provide key competitive advantages to companies and support socially and environmentally beneficial outcomes". See also M. MOZZARELLI, *Digital Compliance: The Case for Algorithmic Transparency*, in S. MANACORDA — F. CENTONZE (a c. di), *Corporate Compliance on a Global Scale*, Cham, Springer International Publishing, 2022, 281.

[95] A recent empirical research by Salesforce on 14.000 workers across 14 counties has revealed that over a quarter of workers are currently using AI at work and 55% of them without a formal approval of their employers or (40%) even a against an explicit ban. (SALESFORCE, *More than Half of Generative AI Adopters Use Unapproved Tools at Work*, novembre 15, 2023.

[96] According to the mentioned research, 64% of workers have passed off the AI's work as their own (SALESFORCE, *More than Half of Generative AI Adopters Use Unapproved Tools at Work* (n. 95)).

algorithm could easily encourage its hidden use, probably without valid strategies (or even a propre training) for managing the risks associated with such use.

12. *AI and Human Decision-Maker: From Human Replacement to Human Enhancement*

If prohibiting AI usage is not a viable solution, then the latter may lie in better regulation of the interactions between the algorithm and the human. Tackling this challenge requires a depth of exploration impossible in this paper. Considering a future comprehensive exploration of the subject, we can still attempt to recapitulate the considerations made to identify three schematic coordinates to initiate future research.

The first consideration arising from the previous analysis is that the legal perspective on the role of AI in strategic decisions calls for a significant shift: the algorithm cannot be managed as a mere product but must be considered as a tool with which humans interact. In this perspective, the algorithm cannot be regarded as a substitute for human decision-making (human replacement) but rather as an enhancer of available resources, especially at the instructional level, for the human (sole) decision-maker (human enhancement)[97]. Though human enhancement and human replacement have often been placed, along with human augmentation, on an evolutionary scale of the AI role[98], the reflections made in the preceding sections advocate a strong discontinuity in terms of regulation between enhancement and replacement.

---

[97] PETRIN, *Corporate Management in the Age of AI* (n. 36), 982, proposes the concept of advisory AI or augmented intelligence, referring "to a combination of artificial and human intelligence, in which Al does not replace human intelligence, but leverages or improves it by, for example, giving information and advice that would otherwise be unavailable or more difficult and time consuming to obtain".

[98] See PETRIN, *Corporate Management in the Age of AI* (n. 36), 980; J. ARMOUR — H. EIDENMULLER, *Self-driving corporations?*, in *Harv. Bus. L. Rev.* 2020, 10, 87 ff.

The regulatory objective of AI as a tool, and therefore human enhancement, is no longer solely to minimise "product" risks but becomes (also and) primarily the interaction between AI and the human decision-maker with the perspective of preserving humans in command.

12.1 *The Ergonomics of the Algorithm*

From what has been discussed so far, the shift in perspective towards interaction implies a particular focus on the integrated decision-making process between the human decision-maker and the algorithm in at least two directions: on the one hand, the actual control by the human decision-maker must not be questioned, and on the other hand, the process should be permeable to currently available legal strategies.

The risk that the human in command might be reduced to a mere formal statement is real: risks such as automation bias have already been highlighted, representing the human overreliance on algorithmic output or the difficulty (or high cost) for the human decision-maker to understand the algorithmic path and the reasons for a particular outcome over another. In order to structure legal precepts in a way that provides the appropriate tools to enable the human decision-maker not only to interpret the algorithmic result but the true centre of the decision-making process. Instead, it is crucial to reconsider the interaction between humans and machines from an interdisciplinary perspective, potentially drawing on behavioural approaches.

We could define this novel approach as the ergonomics of the algorithm. Its aim would be the human-machine interaction so as to avert human mechanisation in favour of genuine human empowerment.

12.2 *A legal strategy driven AI*

Secondly, and probably most critically from a legal perspective, the ergonomics of algorithms cannot be separated from an AI design driven by legal strategies. It is the most significant legal consequence of the regulatory shift from human replacement to human enhancement.

The loss of reliability of the algorithmic result in strategic decisions requires an intervention in the decision-making process in order to include the legal tools necessary to ensure not the exact result but the correct process.

Recently, it has been observed that there still is "a good blueprint" for the integration of the respective human and algorithmic decision-making processes[99]. This task primarily belongs to the legal scholar: the correctness of the decision-making process, unlike the accuracy of the result, does not respond to mathematical logic but to purely legal parameters conveyed through legal strategies.

Designing the algorithm based on corporate legal strategies limits its use but also provides legal legitimacy.

As observed, it would be wrong to assume that all corporate legal strategies can be effectively implemented through a decision-making process algorithm.

Some strategies, in particular those based on collegiality and diversity of a board of directors, appear to be far from the algorithmic logic (at least current logic), and their actual effectiveness within an algorithmic decision-making process should be demonstrated with particular rigour. Otherwise, the use of AI in that circumstance should be excluded.

For strategies that can be implemented using AI, the priority is defining minimum admissibility requirements. A clear definition is essential as it serves as the entry point for algorithms in the realm of strategic corporate decision-making.

---

[99] SOLOVE — MATSUMI, *AI, Algorithms, and Awful Humans* (n. 58), 17.

In our opinion, these requirements serve to make the use of the algorithm legitimate in the context of corporate decisions takin in incomplete environments and must be conceived in such a way that the decision taken with the help of AI is legally equivalent to the decision taken without such assistance. The key regulatory consequence is that the focus should not be on the algorithm and on how to make it equivalent to a (decent) human being (fair, transparent, accurate, ecc.), but on the interaction between human and machine and on how to make it equivalent to an (adequate) interaction between humans.

Compared to an approach focused solely on the algorithm, aimed, for example, at increasing its accuracy or transparency[100], we believe that taking the interaction between humans and machines as a reference and, therefore, focusing on the (possible) conditions of equivalence between the two processes (with and without the algorithm) facilitates the achievement of some significant results. On the one hand, the integrated approach appears better suited to mitigate the risk of human overload and cases of automation bias. On the other hand, comparing the integrated conduct (humans assisted by AI) with solely human conduct allows for a more

---

[100] See C. RUDIN, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence* 2019, 1, 5, 212, who stated, "since the definition of what constitutes a viable explanation is unclear, even strong regulations such as "right to explanation" can be undermined with less-than-satisfactory explanations". In the same way R. GUIDOTTI ET AL., *A survey of methods for explaining black box models*, in *ACM computing surveys (CSUR)* 2018, 51, 5, 36, noted that "one of the most important open problems is that, until now, there is no agreement on what an explanation is. Indeed, some works provide as explanation a set of rules, others a decision tree, others a prototype (especially in the context of images). It is evident that the research activity in this field is not providing yet a sufficient level of importance in the study of a general and common formalism for defining an explanation, identifying which are the properties that an explanation should guarantee, e.g., soundness, completeness, compactness and comprehensibility. Concerning this last property, there is no work that seriously addresses the problem of quantifying the grade of comprehensibility of an explanation for humans, although it is of fundamental importance", while SELE — CHUGUNOVA, *Putting a Human in the Loop* (n. 53), 16, stated that "the simple inclusion of a human in the loop is unlikely to prevent inaccurate predictions based on algorithmic recommendations".

precise definition of the integrated conduct, linking the minimum admissibility criteria to the rule's final goal.

To give an example, and this is perhaps the most significant example when talking about decision-making processes (and one of the first steps of future research[101]), such a perspective would directly impose as a research question the identification of the minimum requirements for the use of AI to activate the Business Judgment Rule and the consequent protection of the decision (human, assisted by AI) from the merits review by courts[102].

The legal answer to this question will, in our opinion, make it possible to provide much more comprehensive and precise legal guidance, not only on the design of the algorithm but on the entire interaction between man and machine, making it possible to weigh up the various elements that make it up (from the literacy of the human user to the possibility of intervening in the algorithmic patterns by inserting company-specific conditions and qualitative assessments, from transparency of the algorithmic outcome to the degree of approximation of the result, etc.).

13. *Conclusions*

The challenge is undoubtedly hard, but we believe it is necessary to address it, especially from a legal point of view. Strategic decisions demonstrate the limitations of the current regulatory framework for managing decisions based solely on AI. This requires a change of perspective that focuses on empowering the human decision-maker.

This change in perspective enables us to establish a strong foundation for holding the human decision-maker accountable rather than the algorithm.

---

[101] One other interesting topic seems to be the relationship between AI use by decision-makers and the independence of the latter.

[102] See on that subject MOSCO, *AI and the Board Within Italian Corporate Law* (n. 65), 93.

It puts the decision-maker in a position to fully understand the algorithm's potential and take full responsibility for its use.

The first operational applications of this paradigm may be limited, but starting here is necessary for the sustainable development of human-algorithm interaction. This interaction should be able to manage any decision, even a strategic one, despite its incompleteness and high discretion.