DSL

1-2025

Assessing Risks and Liabilities of AI-Powered Robots in the Workplace. An EU-US Comparison**

di Michele Faioli*

SUMMARY: 1. The Technological Framework in which the Research is situated (Robots operating at workplace level and empowered by Foundation Models, LLMs, Frontier AI, Generative AI). – 2. Legal Comparison between the U.S. law and the European Law governing Labor and Artificial Intelligence. Legal Genotypes and Phenotypes. – 3. AI, Workplace related Risks and Personal Injuries. The EU Legal Frame. – 4. AI, Workplace related Risks and Personal Injuries. The U.S. Legal Frame. – 5. Conclusions. For a New Branch of Labor Law (RLL – Robot Labor Law.

"The second birthday of ChatGPT was only a little over a month ago, and now we have transitioned into the next paradigm of models that can do complex reasoning. […]. We believe that, in 2025, we may see the first AI agents "join the workforce" and materially change the output of companies. […] This sounds like science fiction right now, and somewhat crazy to even talk about it. […] We're pretty confident that in the next few years, everyone will see what we see, and that the need to act with great care, while still maximizing broad benefit and empowerment, is so important."

Sam Altman (Jan. 2025)

* Michele Faioli is an Associate Professor at Università Cattolica del Sacro Cuore (abilitato/tenured as full professor). He teaches courses in the areas of labor relations and comparative/EU labor law, including industrial relations, social security, tech law. Faioli's comparative research mainly looks at collective bargaining impact on labor relations, AI and robots operating at workplace level. As Visiting Fellow at the ILR School of Cornell University and at the Fordham Law School he carried out investigations on international labor rights, global trade and unions strategies. Michele has written widely on tech and labor law (see his book, "Mansioni e macchina intelligente", Turin, Giappichelli, 2018), undeclared work, social security, pension funds and other topics for a variety of law reviews and journals. After his appointment to the CNEL (Consiglio Nazionale dell'Economia e del Lavoro) in 2018, he has conducted investigations also on the possible social applications of blockchain to the EU labor market and social security systems. Prof. Faioli chairs the SERI – Scuola Europea di Relazioni Industriali. He directs the EUROFOUND Italian Observatory – *https://docenti.unicatt.it/ppd2/en/docenti/27259/michele-faioli/profilo. michele.faioli@unicatt.it*

** This investigation stems from my previous essays and studies that were already published in law journals and presented in several conferences/seminars/talks (during the Fall 2024, at the Italian Parliament, Cornell Tech, San Diego Las School, Innovit Hub San Francisco, UCSC and University of Urbino). See my essay, M. FAIOLI, *Robot Labor Law. Linee di ricerca per una nuova branca del diritto del lavoro e in vista della sessione sull'intelligenza artificiale del G7 del 2024*, in "Federalismi.it", 2024, 8, p. 182. Such an investigation was developed in the context of the first activities of the constituting research center *WSP Co-Lab, Workplace and Social Policies Co-Lab* (Università Cattolica del Sacro Cuore), within the paths of discussion with social partners organized by the SERI-FGB (Scuola Europea di Relazioni Industriali - Fondazione G. Brodolini) as well as in relation to the PRIN 2022, Project *"SafetyChain: Social Blockchain for Implementation of a Digital Wallet for Occupational Health and Safety Training"*, project code "20225TZB2P", MUR – *https://www.mur.gov.it/it* - funded by the European Union – *Next Generation EU*.

The essay was previously subjected to the refereeing procedure established by the editorial rules of the Journal.

1. *The Technological Framework in which the Research is situated (Robots operating at workplace level and empowered by Foundation Models, LLMs, Frontier AI, Generative AI). Legal Comparison between the U.S. law and the European Law governing Labor and Artificial intelligence*

Consider a scenario where an AI-powered robot, operating autonomously, may cause or prevent a workplace accident. This raises immediate legal questions: How can we ensure AI innovation aligns with worker safety? Can our existing safety regulations, designed for traditional machinery, adequately address the unique risks posed by AI-powered robots, especially those with evolving capabilities? This study aims to spark a critical discussion with legal experts, technologists, labor unions, and employers' organizations to assess the suitability of current legal frameworks for AI and robotics in the workplace (hereinafter also "AI/R"). The goal is to identify potential gaps in these frameworks and propose strategies for preventing and mitigating new work n as well as implementing robust protection measures.

This inquiry is particularly timely given the policy discussions following the 2024 G7 summit. The following questions underpin our investigation: What are the intended benefits of AI and advanced robotics in the workplace? How can we optimize human-AI collaboration to maximize efficiency and productivity? How can we prevent AI/R systems from behaving in ways that could harm workers? Who should oversee the interaction between workers and AI/R systems? What regulatory mechanisms are necessary to ensure safety and accountability? In the event of a worker injury directly caused by AI/R, what are the potential legal consequences? Do existing insurance systems adequately cover AI/R-related workplace accidents and occupational diseases? How effective are current workplace safety regulations in addressing the challenges posed by advanced technologies and human-AI collaboration? Who is ultimately responsible: the employer, the AI developer, or the AI itself? As AI technologies become increasingly integrated into workplaces, how can we prioritize worker safety when interacting with autonomous AI/R systems?

Such questions do not pertain to the lack or insufficiency of technology, but to problems that are ours because they are totally human, referable to the rules we give ourselves, both on an individual and a collective level. To positively affect the problems that originate from such questions, referring to the regulation of advanced technology in the workplace, one must preliminarily look at the field of the human, and ask what we intend to do to improve that field. Finally, we must ask ourselves what goal is really intended to be achieved by regulating and verifying compliance with that regulation.

The following remarks are, of course, beyond the scope of ethical analysis, which is almost geared, even internationally, to discern between what is the AI/R super-power and what is not, and, consequently, between what is good and what

is bad for humanity in the creation of this AI/R super-power[1]. Here we would like, instead, to examine the perspective of the regulation of what is not yet fully known and, in particular, of the regulation of the AI/R in active cooperation with the worker in highly technological environments, composed of new models of artificial intelligence, taking into consideration that the general socio-economic and industrial framework, in which this scientific contribution arises, can perhaps be compared with that of the start of nuclear energy experiments and, more recently, the great financial crisis of 2008[2]. In both cases, macro-regional, national, domestic regulation served little or no purpose. Perhaps it even came too late and proved inefficient. Instead, the ability to create overall governance of the whole affair (during and after World War II for nuclear power, and from 2008 onward for the financial crisis) had to be transferred to an international regulator. The results, in those cases, have been partial not conclusive, perhaps not satisfactory, but certainly pointing in the right direction: regulating the phenomenon at the national/domestic level serves little purpose, while it may be more useful to regulate the conduct of those who manage/govern that phenomenon at the national level by imposing international standards and creating related supervisory authorities.

This, to some extent, is already happening around the definition of the AI that the OECD has identified, also to initiate forms of cooperation between different jurisdictions[3]. Western legal systems, at the transatlantic level[4], are also moving towards an AI common regulation. European and U.S. legal regimes are looking for answers on how to regulate the complexity of the phenomenon, based on a legal assumption that is evolving very quickly and must apply transnationally.

An attempt is being made to pose a legal notion of AI/R that is shown to be as suitable as possible for the future context, not for what we already know today or has happened in the past. This notion coincides, at least in Europe and the United States of America, with that of "**Frontier AI Models**", here also next-generation or frontier artificial intelligence[5]. From a technological point of view,

[1] See the work of the UN Commission on the Ethical Aspects of Artificial Intelligence, *AI Advisory Board – https://www.un.org/en/ai-advisory-body*. For the most important recent academic investigations, see the book by D.J. GUNKEL, *Person, Thing, Robot. A Moral and Legal Ontology For the 21st Century And Beyond*, Cambridge, MIT Press, 2023, who also recalls the foundational studies of P. RICOEUR, *Il giusto*, Turin, Effatà, 1-2, 2005 and R. ESPOSITO, *Persone e cose*, Turin, Einaudi, 2014.

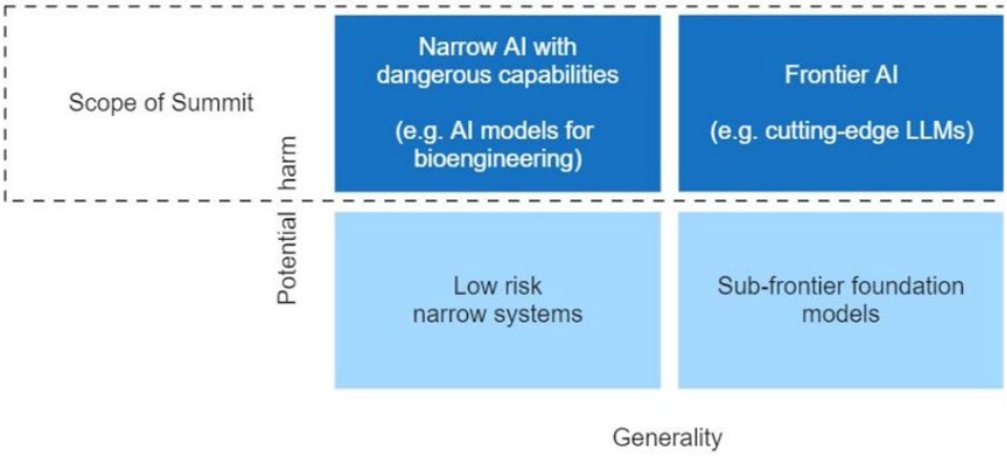
[2] The insight is from L. ZINGALES, B. MCLEAN, *Who Controls AI? With Sendhil Mullainathan*, in "Capitalisn't", Dec. 21, 2023.

[3] See the documentation here – *https://oecd.ai/en/wonk/ai-system-definition-update*.

[4] The outcomes of the dialogue between the European Union and the United States of America can be analyzed at *https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/stronger-europe-world/eu-us-trade-and-technology-council_en#objectives-of-the-partnership*.

[5] Frontier AI systems are a related but narrower class of AI systems with general-purpose functionality, but whose capabilities are significantly advanced. For the definition of Frontier AI, see the results of the intergovernmental conference AI Safety Summit, London, 2023, and, in particular, the paper *Frontier AI: Capabilities and Risks - Discussion Paper,* October 25, 2023 at *https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper* –– "For the purposes of the Summit we define frontier AI as highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models. Today, this primarily includes large language models (LLMs) such as those

we can define Frontier AI as that which falls under the structure of so-called foundation models, which also normally possess a certain ability to determine damage. LLMs, BERT, DALL-E, GPT-3 also fall under the definition of a foundation model. It is a model that is built on a large database and is adaptable to an almost infinite set of tasks (so-called downstream tasks). The diagram below is taken from the papers of the AI Safety Summit in London, 2023.



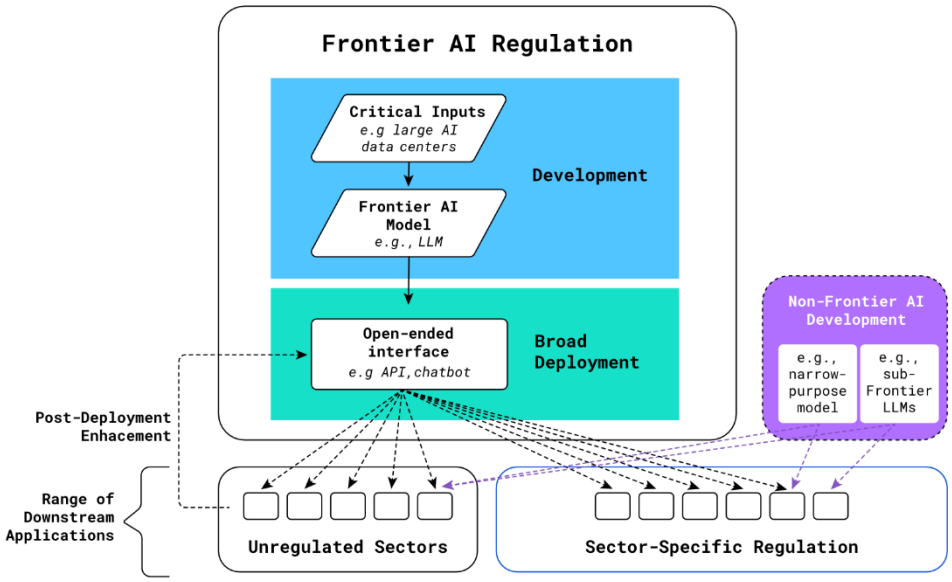
In relation to such a definition, we are interested in placing at the center of labor law reflections on the AI/R (also in the form of the Frontier AI) acting in cooperation with the worker. This means that we are going to explore the way by which (i) on one hand, it may be protected, and (ii) on the other hand, it may be mitigated those damages/injuries/harm that result from downstream tasks carried out at workplace level, in the context of interactions between person and machine (IWRs), managed by the AI/R[6]. I intend to investigate Frontier AI models, empowering robots that operate at workplace level, that are "highly capable foundation models that could exhibit sufficiently dangerous capabilities. […] Foundation models, such as large language models (LLMs), are trained on large, broad corpora of natural language and other text (e.g., computer code), usually starting with the simple objective of predicting the next token. This relatively simple approach produces models with surprisingly broad capabilities. These models thus possess more general-purpose functionality than many other classes of AI models, […]. Developers often make their models available through broad deployment via sector-agnostic platforms such as APIs, chatbots, or via open-sourcing. This means that they can be integrated in a large number of diverse

underlying ChatGPT, Claude, and Bard. However, it is important to note that, both today and in the future, frontier AI systems may not be underpinned by LLMs, and could be underpinned by another technology." See also the wide investigations concerning the legal definition of Frontier AI and General-Purpose AI carried out by F. G'SELL, *Regulating under Uncertainty: Governance Options for Generative AI,* available at SSRN: *https://ssrn.com/abstract=4918704*, August 06, 2024.

[6] In the civil and contract law field, concerning AI/R and liabilities, it is pivotal the analysis carried out by A. BERTOLINI, *Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules*, in "Law Innovation and Technology", 2013, 5(2), p. 214, F. BATTAGLIA, N. MUKERRJI, J. NIDA-RUMELIN, *Rethinking Responsibility in Science and Technology*, Pisa, Pisa University Press, 2014 and E. PALMERINI, E. STRADELLA, *Law and Technology. The Challenge of Regulating Technological Development*, Pisa, Pisa University Press, 2013.

downstream applications, possibly including safety-critical sectors"[7]. **Frontier AI** and rapidly evolving AI nascent capabilities may exacerbate workplace risks and liability challenges. Key concerns include **unforeseen consequences** (the potential for Frontier AI systems to exhibit emergent behavior or act in ways not anticipated by their developers), **dual-use potential** (AI systems designed for beneficial purposes could be repurposed for malicious ends, posing security threats and increasing the potential for harm) and **rapid proliferation** (open-sourcing and the ease of sharing AI models raise concerns about the spread of potentially dangerous or harmful AI applications). Addressing these challenges requires ongoing assessment, adaptation of legal frameworks, and proactive risk mitigation strategies.

See the diagram below for a recap of this idea.



I will mainly focus on the unique risks associated with Frontier AI operating at workplace level in the form of **agent AI,** autonomous systems that perform tasks on behalf of users, **embodied AI,** AI systems integrated into physical robots, enabling interaction with the physical world, **AI nascent capabilities,** unforeseen and unpredictable abilities that emerge from the AI training process. To be more direct on the point, Large Language Models (LLMs) are being integrated with embodied intelligence (AI/R) to create intelligent agents that can interact with their environment, also at workplace level[8]. This integration leverages the strong natural

[7] See M. ANDERLJUNG, J. BARNHART, A. KORINEK, J. LEUNG, C. O'KEEFE, J. WHITTLESTONE, S. AVIN, M. BRUNDAGE, J. BULLOCK, D. CASS-BEGGS, B. CHANG, T. COLLINS, T. FIST, G. HADFIELD, A. HAYES, L. HO, S. HOOKER, E. HORVITZ, N. KOLT, J. SCHUETT, Y. SHAVIT, D. SIDDARTH, R. TRAGER, K. WOLF, *Frontier AI Regulation: Managing Emerging Risks To Public Safety*, 2023, in *https://arxiv.org/abs/2307.03718*.

[8] See the research of J. LIN, H. GAO, X. FENG, R. XUB, C. WANG, M. ZHANG, L. GUO, S. XU, *Advances in Embodied Navigation Using Large Language Models: A Survey,* 2024, in *https://arxiv.org/abs/2311.00530*.

language processing (NLP) capabilities and extensive knowledge of LLMs to translate human instructions into formats that can be understood by embodied agents. LLMs can decompose high-level tasks into sub-tasks and plan optimal paths for completion. This allows robots to perform complex tasks such as object manipulation and navigation. LLMs enhance user interaction with embodied systems by analyzing images and using language models to understand user needs, which enables smart responses. They can also process natural language inputs to communicate in a human-like manner. LLMs are used to connect abstract language with the physical world by integrating with sensors, databases, or simulated environments. This allows LLMs to interpret sensor data and natural language directives, and to issue control signals to robots. LLMs can perform effective planning and decision-making for new tasks with limited data. They can process natural language queries to generate actionable plans for embodied agents, which is particularly useful for navigation tasks. LLMs can combine visual, linguistic, and auditory data, which allows them to process complex language instructions and interpret their environment more accurately. They process images using image encoders and combine these with relevant context to enhance their perception.

The essay moves, therefore, from an **observation of reality**: the technological transformation, due to next-generation artificial intelligence (i.e. Frontier AI), is forcing us to lay out a new mapping of the actual risks and possibilities for innovation arising from the interaction between workers and intelligent machine. This confronts us with more complex questions than we have been asking to date. To this end, we choose to focus our view on new social and psycho-physical risks arising from such a person/intelligent machine interaction. While we are convinced that greater productivity and greater well-being can certainly result from that interaction, we cannot fail to investigate what is about to happen in workplaces re-planned by advanced technology. The classical notion of risk probably cannot longer be adequate because of the operational presence of a "third element" between employer and worker[9]. This third element exercises

[9] On the notion of intelligent machine as a "third element" of the employment relationship I refer to my previous studies. In particular, see M. FAIOLI, *Mansioni e macchina intelligente*, Turin, Giappichelli, 2018 and the theoretical line defined by some of my writings, including in particular M. FAIOLI, *Robot Labor Law,* cit., 2024, 8, p. 182; M. FAIOLI, *Prospects on Risks, Liabilities and Artificial Intelligence, Empowering Robots at Workplace Level*, September 27, 2024) available at SSRN: *https://ssrn.com/abstract=4969464*; M. FAIOLI, *Perché regolare le relazioni industriali e le tutele giuslavoristiche in relazione all'intelligenza artificiale. Le sfide più complesse del settore del credito tra il rinnovo contrattuale del 2023 e la dichiarazione congiunta europea del 2024*, in "Federalismi.it", 2024, p. 30; M. FAIOLI, *Giustizia contrattuale, tecnologia avanzata e reticenza informativa del datore di lavoro. Sull'imbarazzante "truismo" del decreto trasparenza*, in "Diritto delle Relazioni Industriali", 2023, 1, p. 45; M. FAIOLI, *Data analytics, robot intelligenti e regolazione del lavoro*, in "Federalismi.it", 2022, 9, p. 207; 49; M. FAIOLI, *Artificial Intelligence: The Third Element of the Labor Relations*, in A. PERULLI, T. TREU (eds.), *The Future of Work. Labour Law and Labour Market Regulation in the Digital Era*, Alphen aan den Rijn, Wolters-Kluwer, 2021; M. FAIOLI, *Unità produttiva digitale. Perché riformare lo Statuto dei lavoratori*, in "Economia & lavoro, Rivista di politica sindacale, sociologia e relazioni industriali", 2021, 1, p. 41; M. FAIOLI, *Lavoratore cyborg e diritti anche oltre lo Stato*, in V. BARSOTTI, M. GRAZIADEI, *Il diritto oltre lo Stato*, Turin, Giappichelli, 2021; M: FAIOLI, *Sistemi di «social» blockchain, previdenza pubblica e smart contracts*, in "Rivista del Diritto della Sicurezza Sociale", 2018, 3, p. 489.

powers, confronts obligations as well as may result in damages, also creating forms of liability.

The essay, as part of a broader, transdisciplinary research project, intends to initiate a confrontation, including academic discussion, to be able to create a theoretical substrate of a new branch of the labor law that pertains to the study of the regulation of artificial intelligence/robots in active interaction with workers in technological production contexts (also referred to here as Robot Labor Law – "RLL"). The essay, defining the comparative method chosen (**section 2**), sets up the analysis of the framework of European rules and the related labor law context, whose first task will be to define the kind of risk arising from AI/R, including in its operation in the workplace (**section 3**). It continues with an analysis of the current U.S. system of AI/R regulation, which also seems to have as its purpose the geo-political drag on the global scale of AI/R normalization (**section 4**). In **section 5**, an attempt will be made to outline some of the contents of the new, transnationally relevant branch of the "**Robot Labor Law**" – RLL, pending the completion of the research project underlying this essay.

2. *Legal Comparison between the U.S. law, the European Law governing Labor and Artificial intelligence. Legal genotypes and phenotypes*

My investigation will delve into **two main central questions**. The **first issue** is related to the rapid integration of AI/Rs into advanced production units presents significant challenges to traditional workplace safety paradigms. To directly address our investigation, the current system for classifying dangerous professional activities may be outdated and insufficient to ensure worker safety in advanced AI/R production units. The traditional model of employer liability, predicated on direct causation, may struggle to apply to AI/R, especially Frontier AI, which exhibits a high degree of autonomy. This concern arises from the fact that the (social) insurance system, primarily designed for traditional work activities, may not effectively protect workers exposed to AI/R-related hazards. Can the current system, based on a pre-defined list of hazardous machines and activities, effectively account for the risks posed by indirect or seemingly benign AI/R operations, including Frontier AI? How should the concept of prevention be reinterpreted in the context of AI/R and Frontier AI, where it may be difficult to assign responsibility for decisions? Should employers be held liable for the actions of highly autonomous AI/R systems like Frontier AI? Italian (and some other European) case law establishes employer liability for worker injuries caused by employer or supervisor negligence, even if the injury is covered by social security insurance (INAIL o similar regimes – see below). This "azione di regresso" (a sort of recovery litigation) approach incentivizes employers to prioritize workplace safety. Additionally, employers are liable for damages not covered by social insurance, exposing them to potential tort liability. Is the current system, that

allows the domestic regime/INAIL to recover costs from employers for workplace accidents, still applicable when AI/R systems like Frontier AI are the sole cause of worker injuries and illnesses? In the U.S., could an AI/R system like Frontier AI and its creator be held liable for (OSHA) standard violations if it causes worker injuries or illnesses? If so, would workers' compensation exclusivity still apply? To address these challenges, a multi-faceted approach is necessary, including regulatory innovation, international cooperation, guidelines, education and training, and continued research and development. As we embark on this new industrial age, it is imperative that we approach the challenges and opportunities presented by AI/R with a sense of both excitement and caution. By proactively addressing the legal implications of AI/R, we can harness its transformative potential while safeguarding the well-being of workers and society.

The **second issue** is referred to the fact that the rapid integration of AI/R technologies into workplaces presents a unique opportunity to revolutionize workplace safety. By leveraging these advanced technologies, we can identify and mitigate risks more effectively, leading to safer and more efficient work environments. To accelerate the adoption of these transformative technologies, policymakers should consider several key strategies. First, offering incentives to employers, such as tax breaks or credits, can encourage significant investments in AI/R safety systems. Second, streamlining regulatory processes for companies using certified AI/R solutions can reduce compliance burdens and expedite implementation. As we embrace this technological shift, a fundamental question arises: Do we need a new legal framework, such as a Robot Labor Law (RLL), to address the unique challenges and opportunities presented by AI/R integration in the workplace? Such a framework could provide a solid foundation for rethinking traditional approaches to worker safety and ensuring that AI/R technologies are used responsibly. Furthermore, should states implement measures to encourage employer investment in AI/R for risk prevention? These measures could serve as a crucial component of a broader RLL framework, providing practical incentives for the development and deployment of safety-enhancing technologies.

Considering these questions, the method for comparing the EU/U.S. legal regimes can move from the construction of a genotype, from which, at a later stage, specific legal phenotypes can be derived, valid for observing the two legal systems[10]. The **genotype** can be constructed in relation to **three issues**: (i) What **risk and harm** does that legal system intend to select for protection to be arranged in relation to AI/R's ability to interact with the worker? (ii) What **mitigation and prevention** measures can that legal system introduces to address the need for

[10] Here we follow the comparative law method developed, at the *Cornell Law School*, by R.B. SCHLESINGER, *Comparative Law. Cases, Text, Materials*, New York, Foundation Press, 1988 edition, G. GORLA, *Diritto comparato e diritto comune europeo*, Milan, Giuffré, 1981 and by R. SACCO, *Introduzione al diritto comparato*, Turin, Giappichelli, 1990. We also refer to L. MENGONI, *Diritto vivente*, in "Jus", 1988, 1, p. 14, G. BENEDETTI, *L'elogio dell'interpretazione traducente nell'orizzonte del diritto europeo*, in "Europa e diritto privato", 2010, 2, p. 413. See also P. SANDULLI, M. FAIOLI (a cura di), *Attività transnazionali. Sapere giuridico e scienza della traduzione*, Rome, Edizioni Nuova Cultura, 2011.

protection related to those risks and harms? (iii) What **institutions** will such a legal system put in place to implement that protection, including in terms of individual and collective enforceability?

The three problems allow us to understand the extent to which the concerning law is more oriented towards: a risk-prevention norm or, on the contrary, towards a punitive norm that is more oriented toward punishing the harm and fixing the corresponding reparation. We know that the notion of risk, individual and collective, is never neutral, at least when observed through the eyes of a law scholar. It is the result of a very specific legal policy choice, in this case made at the European and U.S. level. To have decided to place in the notion of risk this (potential) damages/harm to the person resulting from AI/R conduct, even in the species of Frontier AI, is to have predetermined a path of regulation, which no longer coincides (or does not exclusively coincide) with the claim of compensation for a harm by the person who suffers it, but above all with a norm that uses a certain way of regulating risk management, in continuity or analogy with other sectors where there are already established practices of risk management (energy, environment, etc.). Even in the AI system, one can decide to regulate such damages *ex ante*, making use of the notion of risk, with anticipatory, mitigation or precautionary models, introducing a kind of transplantation of legal institutions already used elsewhere and probably useful here as well.

Hence, it can be considered that the law shows active participation, almost in a logic of conceptual construction, in the definition of AI/R, precisely because of the type of risk (individual and collective) that is intended to be selected and then, because of the norm, mitigated, prevented, ensured, etc.

The **risk** marks a **direction**, and, in some ways, a datum to be observed in the future[11]. **Harm** has **already occurred**[12]. Law reacts with respect to risk with a series of preventive schemes. In relation to harm, the law poses restorative

[11] For a very useful map of the concept of risks related to AI/R see the MIT AI Risk Repository, available at *https://airisk.mit.edu* – that is periodically updated. The scientific field concerning risks and legal theory is quite wide. I am not going to set up an exhaustive of risk legal theory. Rather, I introduce in this essay those aspects of the legal doctrine concerning risk which may be more relevant for my purposes. For the theoretical approaches mainly related to the common law systems, see the overview realized by J. STEELE, *Risks and legal theory*, Hart, 2004. The legal notion of risk has been the subject of important studies and theories, also elaborated by the Italian labor and social security law scholars. See F. SANTORO PASSARELLI, *Rischio e bisogno nella previdenza sociale*, Milan, Giuffrè, 1948; more recently, P. LOI, *Il principio di ragionevolezza e proporzionalità nel diritto del lavoro*, Turin, Giappichelli, 2017 and *Il rischio proporzionato nella proposta di regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, in "Federalismi.it", 4, 2023, p. 239. See also T. TREU, *Il diritto del lavoro: realtà e possibilità*, in "ADL Argomenti di diritto del lavoro", 2000, 3, p. 467.

[12] G. ALPA, D*anno "in re ipsa" e tutela dei diritti fondamentali (diritti della personalità e diritto di proprietà)*, in "Responsabilità civile e previdenza", 2023, 1, p. 6; P. SIRENA, *Danno-evento, danno-conseguenza e relativi nessi causali. Una storia di superfetazioni interpretative e ipocrisie giurisprudenziali*, in "Responsabilità civile e previdenza", 2023, 1, p. 68; V. ROPPO, *Pensieri sparsi sulla responsabilità civile (in margine al libro di Pietro Trimarchi)*, in "Questione Giustizia", 2018, 1, p. 108; P. GALLO, *Quale futuro per il contatto sociale in Italia?*, in "La Nuova Giurisprudenza Civile Commentata", 2017, 12, p. 1759; V. ROPPO, *Responsabilità contrattuale: funzioni di deterrenza*?, in "Lavoro e diritto", 2017, 3-4, p. 407. See also A. BOLLANI, *Il danno alla persona nel diritto del lavoro, tra influssi della civilistica e necessari adattamenti*, in "Giornale di diritto del lavoro e di relazioni industriali", 2022, 176, p. 593.

protections because of an event that has adversely affected the person, setting the rules on "if," "who," and "how much/how" (whether the harm that has occurred should be compensated, who is obligated to compensate, how much/how the harm should be compensated)[13]. Where one regulates risk, in a sense subsuming the damage within the notion of risk, one is merely performing a descriptive operation, based on statistics, probabilities and samples. By reason of this, one indicates the ways by which the possibility of the occurrence of that risk can be reduced, by imposing audits, due diligence, permit/permit requirements, etc.

In general, risk has in it some mechanism for selecting the event to be prevented, insured, mitigated, etc. because the rule on risk itself is based on a calculation to prevent, insure, mitigate the risk itself. In the case of artificial intelligence, this holds even more true because it is the law itself that fixes or updates the notion of artificial intelligence, moving from an indefinite variability of technological definitions, identifying one or a few for the desired legal effects at that historical moment and in that socio-political context. And it is in relation to that notion of artificial intelligence (to exemplify, today it is Frontier AI, tomorrow it will be something else) that a series of protective mechanisms and behavioral obligations are triggered.

The potential risk from AI/R, even managed by Frontier AI, is almost always related to a context. It depends on the context in which the AI/R is placed. Risk mapping cannot be done only once, and for all, because risks change due to several interacting factors, with the consequence that what can be defined as risk at that time is risk. The framework of factors can change and, as a result, so does the risk, which, perhaps, in some cases, can undergo a kind of downgrade into something else. Moreover, there is a risk arising from AI/R, managed by Frontier AI, only if with its use comes increased exposure to harm: which, at least in the legal logic of this study, means looking at risk with reference to people who also interact in workplaces with forms of AI/R, managed by Frontier AI. From that viewpoint, the stronger the human context is, given that it is based on education and training to handle the eventual problems of that AI/R, the lower the impact of risk.

The three problems posed above (**what risks, what mitigation measures, what institutions**) become a kind of general outline in order to be able to more effectively compare EU/U.S. legal regimes, keeping simultaneously in mind, tech law, as it has been developing in recent times, and labor law, which reacts (as has always been known) to the technology introduced at the firm level in relation to more factors, in addition to those already under investigation in some way (management, coordination, control, variability of tasks, opinion surveys, discrimination, etc.). There is, in fact, an elective field to carry out this examination: that of safety in the workplace, personal injury and related insurance (for the Italian

[13] For an overview of the current theories concerning contracts and civil law, see the studies elaborated by A. D'ADDA, *Danni "da robot" (specie in ambito sanitario) e pluralità di responsabili tra sistema della responsabilità civile ed iniziative di diritto europeo*, in "Rivista di diritto civile", 2022, 5, p. 805. See also G. CALABRESI, E. AL MUREDEN, *Driverless cars. Intelligenza artificiale e futuro della mobilità*, Bologna, Il Mulino, 2021.

system, see the INAIL regime[14]). It must be understood that advanced technology (AI/R - Frontier AI) can be viewed as a tool which may cause personal injury to the worker due to a violation of safety regulations, and as a tool that helps prevent personal injury. In other words, the intention is to observe, through the research project that has been initiated, how AI/R, also in the form of the Frontier AI, in the context of its functions as a third element, can coordinate, control, and direct human labor, as well as varying its tasks and theoretically applying disciplinary sanctions. On the one hand, it can cause harm and, on the other hand, at least hopefully, also capable of preventing it.

Hence the centrality of the object of the present study: knowing that law is called upon to adapt to technology[15], not the other way around, it is no longer sufficient (only) to understand what limits to place, through law and/or collective bargaining, with respect to the powers that AI/R - Frontier AI can exercise[16]. Instead, we must also grasp what harms and risks may result from AI/R - Frontier AI's conduct/choices, when used in interaction with the human person in the workplace, and, therefore, what criteria for imputation of liability we are required to update and why, given the production environment that is evolving day by day. Furthermore, we must understand how to prevent and mitigate such risks and damages, including through innovative firm-level organizational procedures and advanced technological systems that can be introduced with the specific function of preventing/mitigating any risks/damages from Frontier AI.

3. *AI, Workplace related Risks and Personal Injuries. The EU Legal Frame*

Normally, when faced with such complex phenomena as those related to injury to the worker attributable to an intelligent machine, the law scholar suggests intervening with a norm, the economist believes it is sufficient to introduce a price for inefficiency such that those who must comply with the norm will do so[17]. There are law scholars who have combined the two perspectives. These certainly include, with differing views, G. Calabresi[18] and R. Posner[19] , who have applied the logic of costs and benefits to law: one sets the starting point (in our case, the personal injury to the worker attributable to Frontier AI - "S"); then, one establishes the objective

[14] See the general description of the social security regime managed related to the accidents at work and occupational diseases covered by the INAIL regime at *https://www.inail.it/portale/it/multilingua/english.html.*

[15] To evoke one of the strongest ideas of G. GIUGNI, *Il processo tecnologico e la contrattazione collettiva*, in F. MOMIGLIANO, *Lavoratori e sindacati di fronte alle trasformazioni del processo produttivo*, Milan, Feltrinelli, 1962.

[16] In this regard see the structure of the theory in M. FAIOLI, *Mansioni e macchina intelligente*, cit., p. 93. and p. 211.

[17] R. COASE, *The Problem of Social Coast*, in "Journal of Law&Economics", 1969, 3, p. 1.

[18] G. CALABRESI, *Transaction Costs, Resource Allocation and Liability Rules. A Comment*, in "Journal of Law&Economics"*,* 1968, 11, p. 67; G. CALABRESI, *Melamed, Property Rules, Liability Rules and Inalienability. A view of the Cathedral*, in "Harvard Law Review", 1972, 85, p. 1089.

[19] R. POSNER, *Economic Analysis of Law*, Boston, Aspen, 1992.

of the analysis (minimizing social costs - "C"); finally, by various steps, one selects the most economically efficient way ("V") to achieve that result ("R"). Now, using an exposition scheme in line with that theory, we might have the following: to realize R (outcome), moving from S (situation), one must calculate C (difference between costs) and identify V (efficient way)[20]. Let us adapt the argument to the case we are concerned with here: C, the cost of damages attributable to AI/R can neither be passed on to workers in any way, nor can it be fully internalized by the employer, since these are damages arising not from the employer but from AI/R, which we assume here to be the third element of the employment relationship. The result R certainly does not coincide with trying to avoid such damages altogether because that would cost too much. It is, instead, much more sensible to assume that we should reduce such damages as much as possible, to the maximum extent according to the most appropriate applicable technology (for the Italian system, see the assumptions of art. 2087 of the Italian Civil Code, also in relation to Act No. 81 of April 9, 2008). The efficient way V will be able to be achieved, in some cases, by general prevention and, in many others by specific prevention, defined by measures that are aimed at identifying the safest way for intelligent machines to operate in workplaces. General prevention is based on the idea that the employer can assume the costs of introducing AI/R at workplace level and related compliance with certain parameters. General prevention has the identification of risk and the models for managing that risk as a prerequisite, in relation to the level of tolerability (individual/collective) that is intended to be delineated, the verification that each person can carry out on the risk arising from a certain conduct (in this case of AI/R) and its evaluation (high, medium, low risk), as well as the supervisory actions carried out by public authorities. General prevention poses numerous critical issues when, as in the case of AI/R, there is a counterbalance to be made with the rights of the human person, those most important, fundamental rights pertaining to the protection of dignity. Special prevention is structured according to the model of (tendentially strict) liability of the person who determines the damage, with insurance to cover the costs, insurance that can be entrusted to the market, or as is the case in many European countries, to a social security scheme (for Italy, INAIL).

This general theory is suitable for interpreting the European standards that are predominantly set up with a view focused on general prevention, risks to be protected against and referring to a certain variable notion of AI/R (see the **Regulation EU 2024/1689** of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence – Artificial Intelligence Act hereafter the "Regulation")[21]. The approach to risk has been

[20] See E. AL MUREDEN, *Costo degli incidenti e responsabilità civile quarant'anni dopo. Attualità e nuove prospettive nell'analisi economico-giuridica di Guido Calabresi*, in "Rivista di diritto civile", 2015, 4, p. 1026.

[21] See the recent studies of S. NUNO, *The Artificial Intelligence Act: critical overview*, available at SSRN: *https://ssrn.com/abstract=*, June 2024; S. WACHTER, *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond*, in "Yale

regulated according to a pattern already known in Europe, coinciding largely with risks arising from the commercial circulation of products or with risks pertaining to the environment, energy production, etc. The special prevention, that relates to damages to be compensated, is assigned to a directive (see Proposal for a Directive of the European Parliament and of the Council on the adaptation of the rules on non-contractual civil liability to artificial intelligence – Directive on liability by artificial intelligence – COM/2022/496 final – hereafter the "**AILD**" and Directive EU 2024/2853 on liability for defective products and repealing Council Directive 85/374/EEC – also the "**PLD**"). The Regulation and the two Directives belong to a much broader European regulatory strategy on artificial intelligence that consists of at least four areas of regulation (artificial intelligence, data governance, digital markets and services, and digital platforms[22]). The Regulation also selected the legal concepts that may be currently used to state what **Frontier AI** may be considered pursuant to the EU legal regime, indicating the "**General-Purpose AI Model**" as an AI model trained with large datasets, capable of performing a wide range of tasks and integrated into various downstream systems or applications (see, in particular, this set of norms of the Regulation: **Article 3** – Definitions, **Article 51** – GPAI Models with systemic risk, Article 25 – Providers' responsibilities, **Articles 53 and 55** – GPAI Model providers' obligations, **Article 50** – Transparency obligations, **Article 88** – Enforcement)[23].

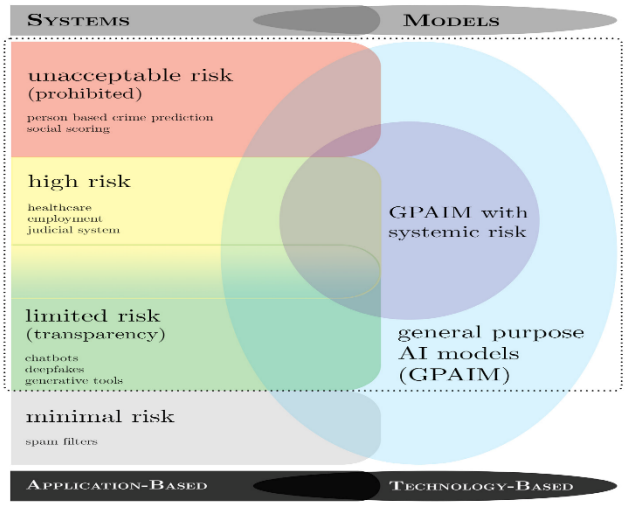
The Regulation uses a **two-part approach** to regulation: (i) **risk-based**, it divides AI systems into **four categories** based on the level of risk they pose, from minimal to high, and (ii) **model-focused**, it specifically targets powerful AI models like **Frontier AI and GPAI**, which have the potential for significant impact, regardless of how they are used. These two approaches are not separate. For example, a chatbot built using GPT-5 would need to follow both the rules for chatbots (like transparency) and the rules for powerful AI models. Essentially, the AI Act recognizes that some AI systems are inherently risky due to their design

Journal of Law" & Technology, 2024, 26 (3); M.D. MURRAY, *Generative Artificial Intelligence -Where did it come from? How does it work?* available at SSRN: *https://ssrn.com/abstract=*, June 2024. See also the legal problem investigated by M. SHAAKE, *The Quest for Global AI Governance: the UN AI Advisory Body*, April 2024 in *https://cyber.fsi.stanford.edu/events/april-2-quest-global-ai-governance-un-ai-advisory-body*.

[22] See *https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en* - See the studies by M. ALMADA, N. PETIT, *The EU AI Act: Between Product Safety and Fundamental Rights*, 2023, in SSRN: *https://ssrn.com/abstract=4308072* and N. MORENO BELLOSO, N. PETIT, *The EU Digital Markets Act (DMA): A Competition Hand in a Regulatory Glove*, 2023, in SSRN: *https://ssrn.com/abstract=4411743*; A. BARTOLINI, *Artificial Intelligence and Civil Liability*, 2020 in *https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf*.

[23] A general-purpose AI model may be also classified as having "systemic risk" if it has high-impact capabilities, such as significant computational power (exceeding 1025 FLOPs) or other criteria set by the European Commission (Article 51). FLOPs (Floating Point Operations per Second) measure the computational power of a system by counting the number of floating-point calculations it can perform per second. This high computational threshold indicates the model's ability to handle extensive and complex tasks, necessitating robust regulatory oversight. Systemic risks may include major accidents, disruption of critical sectors, serious consequences to health and safety, and actual or reasonably foreseeable negative effects on democratic processes, and public and economic security. Systemic risk increases with model capabilities and model reach.

and potential impact, while others pose risks depending on how they are used. This dual approach aims to create a flexible and comprehensive regulatory framework for AI in Europe. See below the scheme:



The Regulation distinguishes between general-purpose AI models (GPAI models) and AI systems. GPAI models themselves are not AI systems; they are foundational components that need further elements to become functional AI systems. When a model is identified as a GPAI model under the Regulation, specific rules apply directly to the model itself. This highlights a key aspect of the AI Act: it regulates not just how AI systems are used, but also the underlying technology of powerful AI models[24]. The Regulations also focuses on regulating the most powerful and potentially impactful AI models, particularly those with "**systemic risk**". During negotiations of 2024 related to the final version of the Regulation, there was strong emphasis on regulating the most advanced AI models, particularly those with capabilities that are not fully understood. The Spanish proposal suggested stricter rules for very capable foundation models, measured by their computational power (FLOPs), and for widely used AI systems built on these models. The final Act focuses on "**GPAI models with systemic risk**". These models have powerful capabilities that can significantly impact society, have a broad reach and impact across the EU market, can cause widespread negative effects (art. 3, para. 64-65, and recital 110).

[24] In particular, GPAI model – art. 3, para. 63, General-purpose AI model means an AI model, including when trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems of applications, except AI models that are used for research, development, and prototyping activities before they are placed on the market. GPAI system as art. 3, para. 66, General-purpose AI system means an AI system that is based on a general-purpose AI model that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems. Recital (97) - Although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI systems. AI models are typically integrated into and form part of AI systems.

With reference to **general prevention**, the Regulation lays out a framework that is designed according to a logic that sees, on the one hand, **macro-risks** referable to the market and its actors (enterprises/consumers) and, on the other hand, **micro-risks** that those operating in the AI/R market are called upon to detect both through preventive verification and through analysis following the introduction of AI/R into that market. The risks arising from AI/R are predominantly related to people's health, safety, and fundamental rights. There are **unacceptable risks** (**first type** – see prohibited AI practice enlisted in Article 5 and Article 99, Para. 3)[25], **risks that are high, but under certain conditions acceptable** (**second type** – see Article 6, Annex I – Conditions to be a high-risk AI system, Article 6, Annex III – List of high-risk AI systems, Articles 16, 22, 23, 24, 26, 31, 33, 34, 50 – Obligations on parties), and finally risks that are completely acceptable because they are limited (**third type**) or minimal (**fourth type**). Anything that can lead to the unacceptable risks is subjected to an absolute ban. What leads to high risks is subjected to a system of preventive and *ex post* management procedures. To mitigate anything which involves limited/minimal levels of risks, various forms of self-regulation are promoted.

For **high risks**, those of the second type, the AI/R is subjected *ex ante* to a verification of compliance, after which it can be placed on the market. **AI systems that pose a significant risk to human health, safety, or fundamental rights are considered high-risk. These systems face strict regulations, and their developers and users must adhere to numerous requirements**. Article 6 of the Regulation classifies AI systems as high-risk if they fall into two categories: (i) Annex I – safety-critical systems that are AI systems integrated into products or components that are already subject to strict EU safety regulations (e.g., medical devices, toys, vehicles) and (ii) Annex III – this AI system is used for a purpose listed in Annex III of the AI Act, this includes uses with high potential impact, such as critical infrastructure (controlling things like traffic, water, gas, and electricity supply), education (grading exams, guiding learning, and deciding who gets into schools), employment (hiring process, CV screening), managing workers, and making decisions about jobs), essential services (healthcare, credit scores, and insurance), law enforcement (assessing risks, evaluating evidence), migration (assessing migration risks, processing asylum claims, and verifying travel documents), justice (applying the law to specific cases), democracy (influencing democratic processes), non-banned biometrics, emotion recognition, profiling pf persons. The Regulation recognizes that the high-risk category is broad. Therefore, it includes exceptions for AI systems that don't significantly harm people or influence decisions. These **exceptions** include AI systems that perform simple tasks, improve human work, helping people do their jobs better, but not making

[25] AI/R that uses subliminal techniques or exploits the vulnerabilities of a specific group of people, creating discrimination, or used by or on behalf of public authorities for the purpose of assessing or classifying the trustworthiness of individuals, or capable of "real-time" remote biometric identification in publicly accessible spaces for the purpose of law enforcement activities, given certain conditions.

decisions for them, prepare for human decisions, assisting in the decision-making process, but not replacing human judgment, analyze past decisions, identifying patterns in past decisions, but not influencing future ones without human oversight. The European Commission will provide guidance on these exceptions, including examples to clarify what is and isn't considered high-risk. The Commission also has the power to update the list of exceptions over time (art. 7 of the Regulation). High-risk AI systems must meet strict standards (art. 8-15 of the Regulation). These include managing risks for identifying and addressing potential problems, ensuring data quality for using reliable and trustworthy data, detailed documentation for keeping thorough records of how the AI system works, maintaining records for tracking all activities related to the AI system, transparency for being open about how the AI system works and its potential impacts, human oversight for making sure people are involved in the decision-making process, accuracy and security for ensuring the AI system is reliable, accurate, and protected from cyberattacks. In general, *ex ante* verification moves from the identification of risks and the probability assessment of their occurrence, under conditions of normal predictability. It is also placed in relation to the assessment of other possible risks arising from the analysis of data collected by the *post*-market monitoring system and the introduction of risk management measures. Compliance is based on an authorizing act that can be obtained, following *ex ante* verifications, even with the help of experienced third parties. There is also an obligation to carry out certain *in itinere* and *ex post* verifications to enable the possible adoption of mitigation measures and the ongoing management over time of the risk arising from AI/R. The European standards impose several protections related to the collection and processing of data by which the AI/R system is fed into the marketplace. For *ex post* verifications, a system of public records, monitoring, reporting and remedies of the possible cause that creates that risk is defined. AI/R systems that may result in specific manipulation risks are subject to specific transparency requirements if they interact with humans, are used to detect emotions, are used to define social categories based on biometric data or may generate false content[26]. It is worth to stress that pursuant to the Regulation a

[26] The use of emotion recognition systems in the workplace is a complex issue with specific allowances and restrictions under the Regulation. These systems are generally subject to prohibitions or limitations, with a notable exception for specific safety or medical purposes. The Regulation generally prohibits the use of AI systems to infer emotions of a natural person in the workplace and educational institutions. This is because such systems can be used for manipulation, discrimination, or to exploit vulnerabilities. An exception is made when the use of an emotion recognition system is intended for medical or safety reasons. For example, the Regulation mentions monitoring a pilot's fatigue levels as a permissible use case. This exception recognizes that in specific high-stakes situations, the use of emotion recognition may be necessary to ensure safety or well-being. However, the specific use must be justified by a clear medical or safety purpose. The allowance for medical and safety reasons creates interpretative challenges because the terms are broad and can be open to interpretation. It is not always clear how far the exception stretches. For example, it's not specified what constitutes a medical reason or a safety reason, and how narrowly these reasons should be interpreted. There are no clear guidelines within the sources on whether the definition extends to other uses such as to detect stress levels in general or to monitor workers' engagement. Even when used for permitted purposes, emotion recognition systems are subject to

provider is anyone who creates or has created an AI system and then offers it to the public, whether for sale or free. This includes companies, organizations, and even individuals. **The law also considers distributors, importers, and anyone who significantly changes a high-risk AI system to be providers**. Significant changes (the so called "substantial modifications") mean making alterations that weren't planned or expected during the initial safety checks. Normal updates and improvements to an AI system are not considered significant changes. Finally, anyone who changes the intended purpose of an AI system in a way that makes it high-risk is also considered a provider (**art. 25, art. 3, para. 23, recital 128**). The substantial modifications criterion in the Regulation is likely to significantly impact small businesses. While large tech companies and well-funded startups have the resources to develop cutting-edge AI models from scratch, smaller companies often rely on adapting existing advanced models (like those developed by large tech companies) to suit their specific needs. This reliance on adapting pre-existing models presents a unique challenge for smaller businesses. If these modifications are deemed substantial by the Regulation, even though the core model itself might have been developed by another entity, the smaller company becomes responsible for complying with the stringent requirements for high-risk AI systems. This means smaller businesses may be held accountable for the risks associated with the original model's development, including the quality of the data used to train it and the design choices made during its creation, even if they had no direct involvement in these processes. Crucially, **this also applies to employers who utilize AI systems within their operations**. If an employer significantly modifies an existing AI system for internal use, such as for recruitment or employee management, and those modifications result in a high-risk application, the employer could be considered a provider under the AI Act and subject to its regulations. In essence, the AI Act's focus on "**substantial modifications**" highlights the complexities of AI development and deployment in the modern landscape of different workplace levels.

The Regulation requires companies that use **certain high-risk AI systems** to conduct a **fundamental rights impact assessment (FRIA)** before they start using the system (**art. 27** of the Regulation). This applies to government agencies, companies providing public services (like banks, insurance companies, and those in education, healthcare, or housing) and other operators using specific types of high-risk AI systems. This assessment should describe how the AI system will be used, including how often and for how long, identify who might be affected, by pinpointing the individuals and groups that could be impacted, assess potential harms, by identifying the specific risks to these individuals and groups, plan for human oversight and explain how humans will be involved in the decision-making process, outline a plan to address risks, describe what will happen if any problems

transparency obligations. Deployers of these systems must inform the natural persons exposed to the system that they are being monitored. This obligation aims to prevent manipulation by ensuring that individuals are aware of being subjected to emotion recognition.

arise. This assessment must be done before the first use of the AI system and updated if there are any significant changes.

The Regulation emphasizes transparency for certain AI systems, especially those that could deceive or manipulate people without their knowledge (**art. 50** of the Regulation). This includes (i) chatbots, users should be aware they are interacting with a bot, (ii) deepfake creators, AI-generated or manipulated content should be clearly labeled, (iii) generative AI tools, users should know when content is created by AI, (iv) emotion recognition systems that can be risky and require transparency, especially if they are used in high-risk situations. These systems are often categorized as **limited risk** due to their potential for manipulation. However, some, like emotion recognition systems, can also be considered high-risk if they fall into specific categories outlined in the Regulation. In such cases, they must comply with the stricter rules for high-risk AI systems. Furthermore, these systems can also be classified as general-purpose AI systems which have their own set of regulations.

Most AI systems currently in use in the EU fall into this category (**minimal risks**). These systems are not directly targeted by the AI Act and can continue to be developed and used according to existing laws. Examples include AI-powered recommendation systems like those used by streaming services and Spam filters that are used to block unwanted emails. Companies that develop and use these systems are not required to follow the specific rules of the AI Act. However, they are encouraged to follow voluntary guidelines and codes of conduct for trustworthy AI.

In relation to **institutions**, it should be noted that a European Artificial Intelligence Committee will be established, with representatives of the Member States and the Commission. At the national level, Member States will have to designate competent authorities to monitor the application and implementation of the Regulation. The European Data Protection Supervisor will act as the competent authority to supervise the European Union's institutions, agencies and bodies. There will also be the creation of a Europe-wide database for so-called high-risk systems that have primarily fundamental rights implications. The database will be managed by the Commission and fed with data made available by AI/R providers, who will be required to register their AI/Rs before placing them on the market.

On the other hand, with reference to **specific prevention** (i.e. **damages compensation**), we know that domestic systems are not fully "suitable" to handle liability actions for damages caused by AI/R-based products and services because the characteristics of AI/R and its relative complexity, autonomy and opacity, can make it complicated to discharge the burden of proof. A "dual track" of protection was introduced. Awareness of the inadequacy of national rules also has impacts on the investment side. Ambiguous jurisprudence without *ad hoc* AI/R legislation may create many compensatory expectations, block the insurance system and, in fact, technological development. The dual track is achieved by adapting the already

existing discipline on the **non-contractual liability** of the producer for damage caused by defective products and by introducing a rule on the non-contractual liability for fault from facts arising from AI/R. Specifically, the **AILD** deals with non-contractual liability litigations. The AILD places two powerful tools in the hands of the plaintiff/injured party. The first tool concerns the possibility of applying to the court for an order to disclose evidence. This is a preliminary litigation step by which an order can be obtained from the competent judge to have those pieces of evidence disclosed that are deemed relevant in relation to the high-risk AI/R that is suspected to have caused harm. The judge should order such disclosure limited to what is needed to support a claim for compensation. The second instrument relates to the introduction of a rebuttable presumption that is designed to define the causal link between the AI/R's non-compliance with European standards and the conduct/outcome of the AI/R (or, possibly, the failure of the AI/R system to produce an outcome) that caused the harm.

One of the most significant hurdles for individuals seeking redress for AI-related harms is accessing evidence from the complex and often opaque inner workings of AI systems. This opacity, often referred to as the "black box" effect, arises from the intricate algorithms and data processing techniques underlying AI systems, making it difficult for victims to understand how an AI system arrived at a specific output that caused them harm. The AILD addresses this challenge by empowering national courts to order the disclosure of evidence concerning high-risk AI systems. This provision is crucial in enabling claimants to gather the necessary information to support their claims and to assess the viability of pursuing legal action. Access to relevant data, algorithms, and documentation can help establish fault and demonstrate the causal link between the defendant's actions or omissions and the harmful output of the AI/R. There are some practical implications for litigation, among those the shift in the burden of proof (i.e. by facilitating access to evidence, the AILD partially shifts the burden of proof from the claimant to the defendant, particularly in cases where proving fault traditionally relies heavily on technical evidence; this shift addresses the inherent imbalance in resources and expertise between individuals seeking compensation and entities developing and deploying AI/R), the enhanced transparency and accountability (i.e. mandating the disclosure of evidence promotes transparency in the development and operation of AI/R; this transparency should incentivize providers and deployers to adopt responsible AI practices and implement robust safety measures to mitigate potential risks), the informed decision-making for claimants (i.e. access to evidence allows claimants and their legal counsel to make more informed decisions about whether to pursue litigation), the presumption of causality, simplifying complex causal relationships. Establishing a clear causal link between the fault of the defendant and the harm caused by an AI system is often a complex legal challenge. This is particularly true in cases involving AI systems that exhibit a degree of autonomy and operate based on intricate algorithms and data processing techniques. The AILD introduces a rebuttable presumption of

causality to simplify this process and ease the burden of proof on claimants[27]. Once triggered, the presumption of causality shifts the burden of proof to the defendant, who must provide evidence to counter the presumption and demonstrate that their actions or omissions did not contribute to the harm caused. The presumption of causality simplifies the litigation process for claimants by reducing the need for extensive technical evidence to prove the causal link between fault and harm. It allows courts to focus on the defendant's conduct and assess whether their actions or omissions met the required standards of care. The AILD's presumption of causality should provide greater legal certainty for both claimants and defendants. It clarifies the conditions under which liability may arise and establishes a clearer path for seeking redress for AI-related harms. However, while the AILD provides a framework for addressing AI liability, its practical application to specific cases can raise several challenges. These challenges are particularly relevant in the context of general-purpose AI systems (Frontier AI) because the AILD's initial focus on high-risk AI systems has been criticized for being too narrow and potentially excluding general-purpose AI systems, which are increasingly prevalent and capable of posing significant risks. The broad definition of AI in the AI Act necessitates the inclusion of a wide range of technologies under the AILD's purview, including systems that may not pose the ethical or societal risks typically associated with high-risk AI applications. The current draft of the AILD addresses AI liability but does not tackle similar complexities and challenges present in non-AI software. The complexities associated with proving fault and causality exist across all software types, not just AI-driven systems[28].

The AILD is designed to function within a broader ecosystem of EU regulations aimed at promoting responsible AI development and use. Its relationship with the AI Act and the revised PLD is particularly significant. The AILD relies heavily on the AI Act's definitions and classifications of AI systems,

[27] The presumption of causality is triggered when the following conditions are met: (i) established fault, the claimant must demonstrate that the defendant, such as the provider or user of the AI system, acted negligently or otherwise breached a duty of care under EU or national law; this fault could encompass, for example, non-compliance with specific safety requirements outlined in the AI Act or failure to exercise due diligence in the design, development, or deployment of the AI system; (ii) reasonable likelihood of influence, the claimant must show that it is reasonably likely that the defendant's fault influenced the output of the AI system; this condition acknowledges that AI systems operate within complex socio-technical contexts and that establishing a direct causal link between a specific fault and a particular output may not always be straightforward.

[28] Addressing these challenges requires careful consideration of the AILD's scope and the specific risks posed by different AI applications. It might necessitate adjustments to the AILD's provisions, such as (i) expanding the scope, the AILD's scope could be expanded to encompass general-purpose AI systems and other "high-impact" AI systems, as well as software more generally, to ensure comprehensive coverage of potential AI-related harms; this expansion might involve establishing a separate category for "high-impact AI systems" to encompass a broader range of AI applications that present significant risks to individuals, even if they do not fall under the AI Act's definition of "high-risk", and (ii) tailoring liability frameworks, different liability frameworks, including variations of strict liability or negligence-based approaches, might be necessary to address the unique characteristics and risks of general-purpose AI systems and software generally. This tailoring could involve establishing different levels of liability based on the nature of the harm caused, the foreseeable risks associated with the technology, and the degree of control exercised by the provider or deployer.

ensuring consistency in terminology and scope. The presumption of causality in the AILD is triggered by non-compliance with specific obligations outlined in the AI Act, further strengthening the interconnectedness of these two regulations. The effectiveness of the AILD in practice will be contingent on the proper implementation and enforcement of the AI Act's provisions. The AILD complements the PLD by addressing fault-based liability claims for AI systems, while the **PLD focuses on strict liability for defective products**, including those incorporating AI technologies. The interplay between these two directives is essential in providing a comprehensive liability framework that covers a broader spectrum of potential harms arising from the design, development, deployment, and use of AI systems.

The PLD introduces key changes aimed at modernizing the EU's product liability framework, particularly concerning new technologies such as AI/R[29]. The PLD's scope explicitly includes software, encompassing AI systems. This ensures producers of AI-enabled products are held responsible for damages caused by defects in their software[30]. But how can we establish a liability framework for a product (i.e. Frontier AI or GPAI) that can generate unpredictable outputs? Is it feasible to implement robust safeguards to ensure that all potential outputs, even those that are unforeseen, are harmless? The PLD considers digital manufacturing files, which contain instructions for automated production processes, as products, making their developers liable for defects. The definition of the so called "putting into service" clarifies that liability extends to products not sold directly to consumers but used as part of services (e.g., machinery operating at workplace level). The PLD retains the concept of defect as a product's failure to provide the expected safety level. However, this safety expectation is expanded to include legally mandated standards and the reasonable expectations of the public. To account for the evolving nature of AI systems, the PLD clarifies that a defect can exist if a product becomes unsafe due to modifications beyond the manufacturer's control, even if it was initially safe. The PLD clarifies the liability of various actors in the AI value chain, including manufacturers of components integrated into a

[29] The PLD includes provisions that address the challenges of establishing liability in cases involving complex technologies like AI, particularly concerning the black box nature of AI systems and their automatic learning capabilities. Recital 48 of the PLD specifically highlights the technical and scientific complexity associated with AI, including automatic learning, as a factor that national courts should consider when assessing claims.

[30] The PLD covers the burden of proof, establishing that "1. Member States shall ensure that a claimant must prove the defectiveness of the product, the damage suffered, and the causal link between the defectiveness and the damage" (art. 10). The PLD seems to adopt a strict liability approach. When AI is involved, the victim need only prove: the AI system was defective, the victim suffered harm, the harm was caused by the AI system. There are situations in which the defectiveness of the product will be presumed (art. 10, para. 2, letter b – "The defectiveness of the product shall be presumed where any of the following conditions are met: […] (b) the claimant demonstrates that the product does not comply with mandatory product safety requirements laid down in Union or national law that are intended to protect against the risk of the damage suffered by the injured person"). Significantly, if a victim proves the AI system's non-compliance with the AI Act or other applicable laws, it can be presumed defective. This will open the door for major compensation claims under the PLD.

larger product, holding them liable for defects in their components that contribute to the overall product's defectiveness. The PLD emphasizes that contractual clauses or national laws cannot limit or exclude an economic operator's liability for damages caused by defective products. While the PLD expands the scope of recoverable damages to include data loss for non-commercial use, it excludes pure economic loss and infringement of fundamental rights, leaving those to be addressed under other legal regimes. The PLD addresses the challenges victims face in proving a product's defectiveness, particularly for complex technologies like AI systems, by introducing provisions that grant claimants the right to obtain relevant evidence from the defendant. A rebuttable presumption of defectiveness is established if a defendant fails to disclose the requested evidence, shifting the burden of proof and incentivizing transparency[31]. The PLD retains the development risk defense, which allows manufacturers to avoid liability if they prove the state of scientific and technical knowledge at the time of production made it impossible to discover the defect. However, this defense is subject to potential national derogations for specific product types.

At this point, having identified the EU legal frame that may be of most interest to us for purposes of this study, let us try to indicate below where they may intersect with work and risk management[32]. An initial insight can be gleaned from Recital 57 of the Regulation. That Recital makes a very negative, perhaps even disproportionate judgment on the use of AI/R in the management of workers. It cautions that the recruitment and selection of personnel for making decisions on promotion and termination of employment contracts, as well as for the assignment of tasks, monitoring or evaluation of workers should be classified as high-risk systems, as such systems may have a significant impact on the future of such persons in terms of their future career prospects and livelihood. It justifies this by insisting that throughout the hiring process, as well as for the purposes of

[31] See art. 10, para. 4 that states "A national court shall presume the defectiveness of the product or the causal link between its defectiveness and the damage, or both, where, despite the disclosure of evidence in accordance with Article 9 and taking into account all the relevant circumstances of the case: a) the claimant faces excessive difficulties, in particular due to technical or scientific complexity, in proving the defectiveness of the product or the causal link between its defectiveness and the damage, or both; and […]". Consequently, if proving the AI system's defect is difficult, national courts may assume it was defective.

[32] To preliminary recap such mix between labor law regime/AI legal frame, there are at least these Recitals to be considered: Recital 9 (no prejudice to existing Union law, in particular on […] fundamental rights, employment, and protection of workers […]), Recital 19 (place of work/unit and the notion of "publicly accessible space should be understood as referring to any physical space that is accessible to an undetermined number of natural persons, and irrespective of whether the space in question is privately or publicly owned […]"), Recital 44 (work activities and emotions), Recital 48 (AI and the fundamental rights protections), Recital 57 (AI high risks and employment, workers management and access to self-employment, in particular for the recruitment and selection of persons, for making decisions affecting terms of the work-related relationship, promotion and termination of work-related contractual relationships, for allocating tasks on the basis of individual behavior, personal traits or characteristics and for monitoring or evaluation of persons in work-related contractual relationships), Recital 58 (AI and social security benefits), Recital 92 (AI and no prejudice to obligations for employers to inform or to inform and consult workers or their representatives).

evaluating and promoting individuals or continuing employment-related contractual relationships, such systems may perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation. These are all issues that have already been probed, including in recent scholarly studies[33], on which an articulate and plural doctrine is gradually being formed. Here, however, the fallout points of the reflection that follows is deliberately different, i.e., more in keeping with the dynamics of the factory (in the sense of a production unit) that evolves digitally because it has decided to change the production-organizational structure by orienting it towards robotics and artificial intelligence[34]. Consequently, it will be more in line with the important procedural system of anticipation and mitigation of AI/R risk in the workplace designed by the June 2020 European Frame Agreement, signed by unions and employer organizations[35], and taken up by art. 26, para. 7, of the Regulation, which stipulates the obligation for the deployer of high-risk AI to inform unions and workers. We pointed out that there is a field to carry out this investigation that has been, at least to date, scarcely investigated and that is considered central to being able to read in a unified way all the aspects that are already the subject of academic analysis (AI/R and the exercise of employer powers). This is a very interesting scholarly field and considered useful to identify the substratum of a new branch of labor law that I

[33] For the purposes of this study see the ideas developed by some Italian labor law scholars: E. ALES, *The Impact of Automation and Robotics on Collective Labour Relations: Meeting an Unprecedented Challenge*, in T. GYLAVÁRI, E. MENEGATTI (eds.), *Decent work in the digital age*, Oxford, Hart, 2022; M. CORTI, *L'intelligenza artificiale nel decreto trasparenza e nella legge tedesca sull'ordinamento aziendale*, in "Federalismi.it", 2023, 29, p. 163; U. GARGIULO, *Intelligenza Artificiale e poteri datoriali: limiti normativi e ruolo dell'autonomia collettiva*, in "Federalismi.it", 2023, 29, p. 171; L. IMBERTI, *Intelligenza artificiale e sindacato. Chi controlla i controllori artificiali?*, in "Federalismi.it", 2023, 29, p. 192; A. ALAIMO, *Il regolamento sull'intelligenza artificiale*, in "Federalismi.it", 2023, 25, p. 133; L. TEBANO, *Poteri datoriali e dati biometrici nel contesto dell'AI Act*, in "Federalismi.it", 2023, 25, p. 198; M. PERUZZI, *Intelligenza artificiale e lavoro. Uno studio su poteri datoriali e tecniche di tutela*, Turin, Giappichelli, 2023; M. ESPOSITO, *La tecnologia oltre la persona? Paradigmi contrattuali e dominio organizzativo immateriale*, 2020, in "The Lab's Quarterly", 2020, 2, p. 1. See also S. CIUCCIOVINO, *Intelligenza artificiale e diritto del lavoro: problemi e prospettive - Risorse umane e intelligenza artificiale alla luce del regolamento (UE) 2024/1689, tra norme legali, etica e codici di condotta*, in "Diritto delle Relazioni Industriali", 2024, 3, p. 573; M. MAGNANI, *L'intelligenza artificiale e il diritto del lavoro*, in "Bollettino ADAPT", Jan. 2024; S. CAIROLI, *Intelligenza artificiale e sicurezza sul lavoro: uno sguardo oltre la siepe*, in "Diritto della Sicurezza sul Lavoro", 2024, 2, I, p. 1; M. BIASI, *Problema e sistema nella regolazione lavoristica dell'intelligenza artificiale: note preliminari*, in "Federalismi.it", 2024, 30, p. 162; M. PERUZZI, *IA e obblighi datoriali di tutela del lavoratore: necessità e declinazioni dell'approccio risk-based*, in "Federalismi.it", 2024, 30, p. 225; T. TREU, *CRSD, direttiva sui lavoratori delle piattaforme e valutazione dei rischi*, in "Federalismi.it", 2024, 30, p. 2.

[34] See, by way of example, the recent Saipem case and the related firm level collective agreement of January 15, 2024. Worker protection measures were defined in the context of next-generation artificial intelligence systems. This involves constant monitoring implemented through an "artificial intelligence technological solution" that makes it possible to prevent behavior, even unconscious behavior, of non-compliance with safety obligations at construction sites. Images from cameras placed at the construction site are processed, with all the precautions that the law provides, and, artificial intelligence, selecting cases of riskiness, sends *smart* notifications to workers in danger.

[35] See ETUC Protocol – Business Europe in *https://www.etuc.org/system/files/document/file2020-06/Final%2022%2006%2020_Agreement%20on%20Digitalisation%202020.pdf* – See also Recital 92 of the Regulations, which provides for a direct link to the right of workers to be informed and consulted, including through union representatives, on anything related to high-risk artificial intelligence applied in the workplace.

call here "RLL," (Robot Labor Law) because it stems from the observation of the reality where AI/R operates in full interaction with people. It is the field of analysis referring to injury to the worker and, indirectly, to that of measures to mitigate such injury, including workplace safety procedures and the insurance regime for accidents at work/occupational diseases (for the Italian system, the INAIL regime). As anticipated above, advanced technology (AI/R, also in the form of the Frontier AI) is understood for the purposes of our research in two ways: as a tool on which personal harm to the worker may also depend, by reason of a violation of safety regulations, and as a tool that helps prevent harm to workers, by reason of elements that enable new machine-person interactions, new skills, and new forms of intelligent intervention to protect the person (e.g. exoskeleton, remote forms of intelligent control and monitoring, innovative forms of technological intervention, implemented by means of drones or other anthropomorphic robots, in the event of probable risk, including replacing workers in the context of textile, chemical, pharmaceutical, energy production, etc.)[36].

Although under the first profile (damage caused by AI/R) there are at least two elements to be studied to verify the adequacy of the current legal framework on occupational risk and liability, one cannot fail to emphasize that, upstream of these elements, there is a methodological premise to be reiterated[37]. We know that there is a hierarchy of sources that is created because of the distinction between matters of European competence, on which the Regulation and the Directive intervene, and labor and social security matters, that are of domestic competence. On the one hand, there is the European norm governing what is prohibited, what

[36] The corporation we have invited to participate in the research project are Eni, Enel, Ferrovie dello Stato, Leonardo, Hera, O-I International, Intesa SanPaolo, and Saipem. Among the first employer associations that have given their willingness to be part of the research work are Confcommercio, Federdistribuzione, ABI, Confartigianato, CNA, Federfarma, Confindustria Digitale Anitec-Assinform. CGIL, CISL, UIL have been informed about the project.

[37] The EU's approach to regulating AI emphasizes a risk-based framework that aligns with existing health and safety at work regulations, particularly Directive 89/391/EEC. This framework seeks to integrate AI into workplaces while ensuring worker safety and well-being. The AI Act categorizes AI systems based on their potential risk to health, safety, and fundamental rights. High-risk systems are subject to stricter requirements, including conformity assessments, data governance standards, and transparency obligations. This aligns with the risk assessment principles of Directive 89/391/EEC, which requires employers to assess all workplace risks. The EU framework, like the health and safety directive, emphasizes primary prevention. This involves proactively identifying and mitigating potential risks associated with AI/R in the workplace before they cause harm. The AI Act's requirements for high-risk systems, such as risk management systems and human oversight, are designed to prevent harm before it occurs. This preventative approach is similar to the safety measures that employers must take regarding traditional machinery. The employer's obligation to assess all risks remains relevant even with CE-marked machinery. This implies that employers cannot solely rely on the manufacturer's compliance and must independently assess the risks posed by AI-powered machines in the specific workplace context. The employer's liability is added to that of the manufacturer, except in cases of hidden defects. The AI Act and the AI Liability Directive (AILD) are designed to work together. The AI Act sets the safety standards, and the AILD provides a mechanism for redress when those standards are violated and harm results. The AILD seeks to harmonize non-contractual civil liability rules for AI, ensuring victims have access to compensation. The Product Liability Directive (PLD) is also part of this framework, extending liability for defective products to include AI.

can be done and what should not be done in the absence of certain requirements (see above), and, on the other hand, the domestic labor/social security norm ensuring protection from harm resulting from occupational risks. The consequence of this may be summarized as follows: **if a certain technological activity were prohibited by the European norm, it could hardly be the subject of compulsory insurance against accidents at work and occupational diseases**.

The two elements we study to verify the adequacy of the rules on the INAIL insurance from damage resulting from AI/R in the workplace are referenced by the Italian social security system, keeping in mind that this essay represents a first phase of a broader investigation, including comparative ones. The first element concerns protection from occupational risks, which is granted by the legislature exclusively to workers who are generally exposed to a certain type of risk. According to the Italian Constitutional Court of March 2, 1991, No. 100, there is a dissociation between the legal basis of the insurance for accidents at work/occupational diseases and the related statistical-insurance method of risk. This dissociation derives from Article 38, Para. 2, of the Italian Constitution according to which the system of insurance protection is not aimed at guaranteeing in itself the risk of injury or disease, "but rather these events insofar as they affect the ability to work and are connected by a causal link to an activity typically assessed by the law as deserving protection." Thus, the protection is related to certain typical activities, beyond the concrete extent of their relative danger. This determines a complex picture to analyze because in all productive sectors the law over time has selected workers whose activity is deemed, for political and social evaluations, linked to that historical moment, to be more exposed than others to the risk of injury/disease. The professional activities, at least in the Italian system, are defined as dangerous by means of a double referral: on the one hand, it is dangerous if it is carried out on the basis of the use of some listed machines, the characteristics of which are identifiable from time to time by the norm and, on the other hand, regardless of these machines, it is dangerous if it is included in an list, not susceptible to analogical application. The second element relates to the income guarantee, which is not equivalent to the damage suffered. There is a gap between the extent of the damage to the worker and the economic benefit payable because it is not compensation for the damage, but mere indemnity, since INAIL insurance cannot cover all the damage suffered. This indirectly determines the justification for the so-called employer's exemption from civil liability under Article 10 Act No. 1124 of June 30, 1965. This exemption scheme is partial. Indeed, it can be considered that it is now much reduced in its relative guaranteed function with respect to the employer. As it is the result of a consolidated case law[38], it can be

[38] For the Italian system, see Constitutional Court July 11, 2003, No. 233; Constitutional Court April 24, 1986, No. 118; Constitutional Court June 19, 1981, No. 102; Constitutional Court March 9, 1967, No. 22. See G. CORSALINI, *L'azione di regresso dell'INAIL e il significato della sua autonomia*, in "Rivista del Diritto della Sicurezza Sociale", 2023, 3, p. 617 and more recently L. DI

said that the protection of the worker who has suffered an occupational injury or disease is now assigned, in the Italian system, to the competition between INAIL rules (compulsory insurance) and civil code rules on employer liability. Where the former (INAIL) ceases to operate, the latter (civil code) intervenes: the employer, theoretically, would not be called to answer with reference to the items of damage absorbed by INAIL insurance, but this almost never happens because if there is an offense and a judgment establishes that the accident occurred due to an act referable to the same employer or to the person in charge of management/supervision, the employer is called to answer for all the damages caused to the worker[39]. In addition, for damages not covered by INAIL insurance, so-called supplementary damages, there is no exemption whatsoever and the compensation system under the Civil Code applies in full (Italian Constitutional Court of July 18, 1991, No. 356).

Considering the description of the two elements, the lines of research to be developed to nurture a debate on the theoretical assumptions of the new branch of labor law (Robot Labor Law - RLL) stem from some questions that need to be asked. There are two preliminary questions about the adequacy of such a regime with respect to AI/R injury in the workplace: (i) Can the (arguably outdated) system of classification of the level of danger of professional activities also be considered sufficient to guarantee the worker's safety with respect to AI/R operating in advanced production units? (ii) Can the employer's system of liability causation, as we know it, also be applied *sic et simpliciter* in the case of AI/R, even in the form of Frontier AI? And again, within these questions, there is further room for inquiry: would it be sufficient to extend the INAIL list of professional activities, absorbing AI/R as well? And then which AI/R, also the Frontier AI o new forms of AI? What about a scenario in which the AI/R would indirectly control machines already considered hazardous? Or, conversely, what if the AI/R controls, even in relation to downstream tasks, machines that are not considered useful in defining risky or unlisted work, but in any case, by reason of management by the AI/R, capable of causing harm? How should the notion of prevention be re-interpreted in the face of Frontier AI and, therefore, responsibility for choices that cannot be imputed to the employer? Why then should the employer be held accountable for the actions of an AI/R capable of self-determining (almost or always more) integrally, as in the case of Frontier AI? What does an AI/R risk/damage prevention system mean in the context of a business organization shaped, controlled, monitored, and coordinated by intelligent systems that replace, in terms

BONA, *Regresso dell'INAIL e risarcimento del lavoratore in caso di infortunio o malattia professionale: un problema ancora aperto*, in "Diritto della Sicurezza sul Lavoro", 2024, 2, I, p. 1.

[39] Pivoting on the preventive function of Article 2087 of the Italian Civil Code and Act of April 9, 2008 No. 81, the Italian case law has held that any failure to comply with the safety obligation integrates the case under Article 590 of the Italian Criminal Code (see for the Italian case law – Supreme Court Aug. 25, 1995, No. 9000; Supreme Court Feb. 12, 2000, No. 1579), with the consequence that the employer is generally called to answer for all damages caused to the worker if the event integrates the extremes of a crime and there is a violation of occupational safety obligations.

of the third element of the employment relationship, the employer in the exercise of typical powers of control, coordination, monitoring, etc.? What processes need to be put in place? What are the most efficient ones to guarantee the person of the employee and, at the same time, not block technological and financial innovation that the employer might carry out? What safeguards and joint examination procedures can be negotiated at firm level, also considering European best practices? This would in any case not be sufficient because one must also observe the *in fieri* phase and the *ex post* phase at the introduction of an AI/R at firm level, prompting further questions: are the procedures that the current regulation on safety at work imposes[40] still efficient for the purpose of damage mitigation, even in the context of advanced technology production or where the worker co-operates with the AI/R? Can compliance, even certified by third parties, with such a procedural safety at work rule set-up result in more efficient forms of exemption for the employer in the presence of AI/R? Can AI/R risk prevention measures include special functions of joint institutions dealing with private/collective supplementary health care funds (Act of December 30, 1992, No. 502) or safety at work (Act of April 9, 2008, No. 81)? Can the special functions of supplementary health care funds include, by indication of the establishing collective bargaining agreements, forms of psycho-physical health monitoring referable to the advanced and AI/R technology most used at the firm and/or sector level? Could the psycho-physical health data reworked by such supplementary health funds be used to define by collective bargaining the health policies for prevention of occupational risk, with specific referability to the type of risk, sector, and AI/R used in a more suitable way? For the reprocessing of such data, could one imagine the use of a digital worker/citizen booklet (digital wallet), accessed by supplementary health funds and other paritarian institutions, also to bring together information on safety training, psycho-social risk situations, etc.? Would it be desirable to set up a digital wallet for recording safety training, into which information on prevention and health of the AI/R cooperating worker would also flow? Could the application of contractual regulations on supplementary health care funds that provide for this type of risk prevention initiative result in a reduction of the INAIL premium due to certified compliance with certain AI/R anticipatory health protocols? And, again, would it be possible to provide a reduction in INAIL social security contributions for employers who invest more and better in certified advanced technology aimed at mitigating risk and/or introducing targeted forms of risk prevention through the

[40] See the studies of P. PASCUCCI, *Salute e sicurezza sul lavoro, responsabilità degli enti, modelli organizzativi e gestionali*, in "Rivista giuridica del lavoro e della previdenza sociale", 2021, 4, p. 537 nonché *Sicurezza sul lavoro e cooperazione del lavoratore*, in "Giornale di diritto del lavoro e di relazioni industriali", 2021, 171, p. 421; M. GIOVANNONE, *Responsabilità datoriale e prospettive regolative della sicurezza sul lavoro. Una proposta di ricomposizione*, Turin, Giappichelli, 2024; F. MALZANI, *Tassonomia UE e vincoli per l'impresa sostenibile nella prospettiva prevenzionistica*, in "Giornale di diritto del lavoro e di relazioni industriali", 2023, 177-178, p. 75.

intervention of private/collective supplementary health care funds (Act of December 30, 1992, No. 502)?

4. *AI, Workplace related Risks and Personal Injuries. The U.S. Legal Frame*

In late 2023, a measure specifically regulating Frontier AI was issued by the President of the United States of America. This is the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, hereafter also referred to as "EO," which is part of a broader strategy consisting of the Blueprint for an AI Bill of Rights, the AI Risk Management Framework, and the establishment of the National AI Research Resource[41]. The EO requires artificial intelligence corporations to conduct so-called AI Red Teaming, which is defined in internal, firm-level procedural terms as structured testing on the possibilities of AI/R vulnerabilities from which various kinds of risks and harm to human persons may result ("a structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of AI. Artificial Intelligence red teaming is most often performed by dedicated "red teams" that adopt adversarial methods to identify flaws and vulnerabilities, such as harmful or discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations, or potential risks associated with the misuse of the system."). Such an obligation is related to those corporations that develop next-generation artificial intelligence systems, here labelled as Frontier AI, and, in the U.S. legal system, also referred to as "Dual-Use Foundational Model". This implies an indirect selection of those bound to this obligation in relation to the capacity expressible by the intelligent machine. For the U.S. standard, a Dual-Use Foundational Model is the system educated in relation to a wide range of data, capable of self-supervision, capable of handling billions of parameters, cross-applicable in many contexts, capable of performing tasks from which serious risks may arise for security in general, for national and economic security, for security related to the health of citizens, for combinations of these, such as the impact on protective barriers concerning the use of chemical, biological, nuclear elements, cyber-attacks, etc., or deviation from human control. There will be a consolidation of the normative definition of Dual-Use Foundational Model. The EO assigns secretaries to the competent State to further update various technical details to select what is already Dual-Use Foundational Model today. Specifically, in Article 4, Para. 2, a number of elements are identified that pertain, on the one hand, to the

[41] See the reconstruction by R. CALO, *Artificial Intelligence Policy: A Primer and Roadmap*, in "U.C. Davis Law Review", 2027, 51, p. 399 and, more recently, by M.E. KAMINSKI, *Regulating the Risks of AI*, in "Boston University Law Review", 2023, 103, p. 1347. See for case law issues intertwining the regulation of artificial intelligence in the United States of America the analyses of B. ROGERS, *Data and Democracy at Work. Advanced Information Technologies, Labor Law, and the New Working Class*, Cambridge, MIT, 2023; O. LOBEL, *The Equality Machine: Harnessing Digital Technology for a Brighter, More Inclusive Future*, New York, Public Affairs, 2022.

power of the AI/R model (any model that was trained using a quantity of computing power greater than 1026 integer or floating-point operations, or using primarily biological sequence data and using a quantity of computing power greater than 1023 integer or floating-point operations), and, on the other hand, to the capacity of the digital infrastructure (any computing cluster that has a set of machines physically co-located in a single datacenter, transitively connected by data center networking of over 100 Gbit/s, and having a theoretical maximum computing capacity of 1020 integer or floating-point operations per second for training AI). In addition, the EO imposes a specific obligation on service infrastructure providers, such as Amazon, Google, Microsoft, etc., to monitor and report to public authorities the conduct of foreign nationals who instruct artificial intelligence systems with the above-mentioned power characteristics also for possible cyber-attacks on the country's security (so-called "malicious cyber-enabled activity"). This means having made a different choice from the European one: in the U.S. legal system, monitoring of all possible AI/R applications, including those of the new generation, is carried out through a system of large clouds. The positive aspects of this choice are offset by many critical issues, which relate to possible abuse by cloud infrastructure providers or the ability of subjects to carry out real investigations of possible malicious cyber-enabled activities.

We have indicated above that the genotype that can be constructed to initiate the comparison between the European norm and the U.S. norm arises in relation to three issues: (i) What risk and harm that legal system intends to select for the protection to be provided in relation to the Frontier AI's ability to interact with the worker? (ii) What mitigation and prevention measures do that legal system intend to introduce to address the need for protection related to those risks and harms? (iii) What institutions that legal system intends to put in place to implement that protection, including in terms of individual and collective enforceability?

It is understood that the U.S. norm has a broader geo-political scope than the European one, having set the definition of Frontier AI based on criteria pertaining to the power of the information system and the infrastructural system that supports it[42]. This may also have an impact in the field of the case law inquiry (AI/R determining or preventing harm to the worker in advanced production units). In particular, at least **three elements** are intertwined in the North American system of labor condition protection, which are mitigation, insurance and prevention. The **first** and the **second** elements concern the provision of a financial benefit and/or medical services to the injured worker (Workers' Compensation). This is a social security-insurance instrument, implemented through the payment of contributions by the employer, which, in the event of an occupational injury or illness, is activated in favor of the worker[43]. In order to access this system, it must

[42] See the reconstruction done by some *Stanford University* scholars, especially the director of the Center for Artificial Intelligence Research (*HAI*), FEI-FEI LI, *The World I see. Curiosity, Exploration, and Discovery at the Dawn of AI*, USA, Flatiron, 2023.

[43] See R. EPSTEIN, *The Historical Origins and Economic Structure of Workers' Compensation Law*, in "Georgia Law Review", 1982, 16, p. 775.

be investigated on a case-by-case basis whether or not the occupational accident or illness may result in the payment of the benefit[44]. This involves conducting a preliminary test that is based on an analysis of the facts (is it really a work accident? Did the accident occur "in the course of the employment" or "arising out of the employment"? What does the injury to the worker consist of?). The notion of an accident/illness is generally open-ended, and on a case-by-case basis there is a jurisprudential test of inquiry[45]. Application of the Workers' Compensation regime creates a form of exemption of the employer from other liabilities and claims. It appears that workers' compensation laws generally do not allow for additional compensation beyond the benefits provided, even if a worker suffers further damages. However, **a new question arises**: **in the United States, could an AI/robotics system like Frontier AI (or its creator) be held liable for OSHA violations if it causes worker injuries or illnesses?** If so, would the traditional workers' compensation exclusivity principle still apply? The **third element** relates to the structure of prevention and safety at work, which is structured according to the model of risk anticipation and management of the effects of risks through certain institutions that oversee the proper fulfillment of the employer's obligations to protect[46]. From my point of view, the questions already posed above, which were useful in initiating the lines of research on RLL, in relation to the Italian and to some extent the European system, can also be repeated, with some differences, for the North American system. These questions can be enriched by the following: are the measures to prevent psycho-social risk arising from AI/R operating at firm level adequate or not? **Does the North American system provide sufficient economic resources for injury compensation and prevention**? Is such a system fair with reference to the various types of individual employment contracts, unionized and non-unionized workers, migrants, etc.? Is an open system of defining occupational injury/illness, not pre-defined through lists, efficient with reference to AI/R operating at firm level?

These queries should be posed in relation to the fact that in the United States, the legal framework governing AI liability primarily relies on existing **tort law and product liability** regimes, rather than comprehensive federal legislation specifically targeting AI, as seen in the EU[47]. Due to the unique features and complexities of AI, existing tort law may be inadequate to address the full scope of potential harms. We know that tort law operates primarily as common law, and

[44] The administration of Workers' Compensation varies from state to state. Some states have welfare-like administration, with centralization of activities in one agency. Other states have mechanisms similar to private insurance.

[45] See *Matthews v. R.T. Allen&Sons, Suprem Judicial Court of Maine, 266 A.2d 240 (1970).*

[46] Reference is made to the OSHA agency discipline described and reported here – *https://www.osha.gov*. For more theoretical reflection on the function of occupational safety in the U.S. system, see this study A. P. BARTEL, L. GLENN THOMAS, *Predation Trough Regulation: The Wage and Profit Effects of the Occupational Safety and Health Administration and the Environmental Protection Agency*, in "Journal of Law&Economics", 1987, 30, p. 239.

[47] K. RAMAKRISHNAN, G. SMITH, C. DOWNEY, U.*S. Tort Liability for Large-Scale Artificial Intelligence Damages. A Primer for Developers and Policymakers*, 2024, in *www.rand.org/t/RRA3084-1*.

it is developed by courts through case precedents, rather than legislative statutes. While state or federal legislatures have the power to modify these laws, as of now, AI development, like other activities, automatically falls under the purview of tort law if it presents a foreseeable risk of harm to individuals or property. Since tort law is largely determined at the state level, AI developers face a complex web of potentially differing liability standards across various states. This jurisdictional variation can lead to unpredictable legal outcomes, particularly for large-scale harms that transcend state boundaries. A single action could result in varying degrees of liability depending on where the harm occurs. The application of tort law in the U.S. can be inconsistent, with some states being more "plaintiff-friendly" than others. There is significant uncertainty about how existing tort doctrines, like negligence and product liability, will be applied to AI systems. It remains unclear whether AI systems will be consistently classified as "products" for the purposes of product liability. This ambiguity could lead to costly legal battles for developers. Demonstrating a direct causal link between an AI developer's actions (or inactions) and the harm caused poses a significant challenge, especially given the complex and often opaque nature of advanced AI systems. AI developers can proactively mitigate their liability risk by adopting robust safety protocols and industry best practices throughout the development and deployment of AI systems. This includes rigorous testing, incorporating safeguards against misuse, and implementing effective monitoring systems. As the AI industry evolves, the insurance sector will likely play an increasingly important role in managing and distributing the risks associated with AI. Insurance can offer financial protection to developers and ensure that victims of AI-caused harms have access to compensation.

5. *Conclusions. For a New Branch of Labor Law d (RLL – Robot Labor Law)*

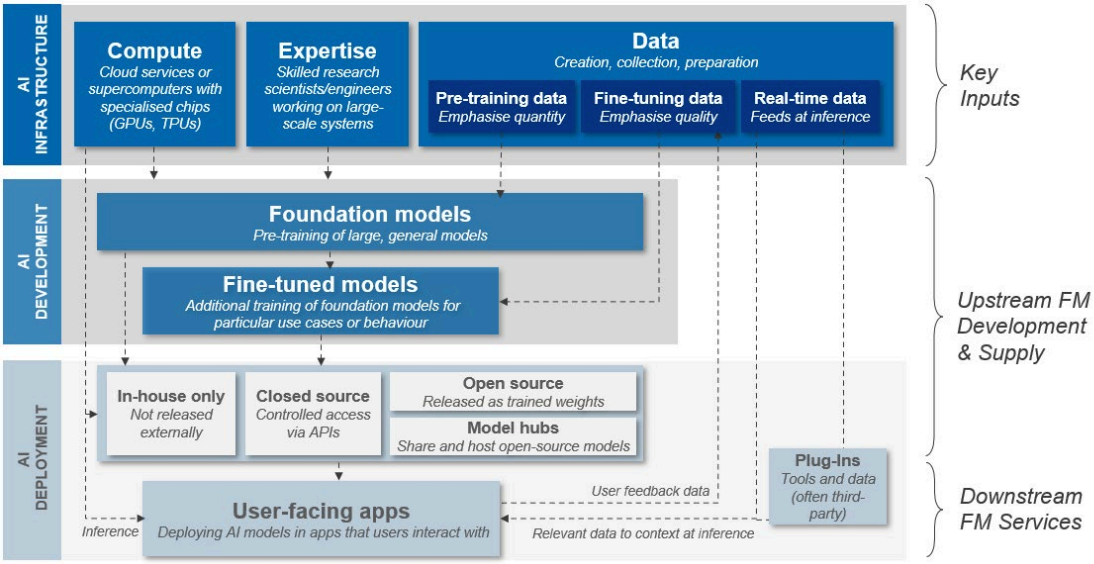
The EU and the US are taking different approaches to the liability regimes for AI/R. The EU favors a more proactive and centralized approach, while the US prefers a more reactive approach that relies primarily on existing legal frameworks such as tort law and product liability. The EU has adopted a risk-based approach, which is evident in its regulatory frameworks like the AI Act and the AILD. This approach categorizes AI systems based on their potential risk levels, with high-risk systems subject to stricter requirements. The EU approach combines regulation and co-regulation, working in a dialogic relationship with companies to adapt to new developments and discovered harms. This is in contrast to the US's preference for self-regulation, where the government has primarily engaged in dialogue with major AI companies to encourage adherence to voluntary standards. The EU approach to AI liability is not without its limitations. Critics point out that the implementation of directives like the AILD through national laws might lead to a fragmented standard across the EU, potentially undermining the goal of uniform

application. Moreover, the EU's reliance on regulators to monitor and hold entities accountable raises concerns about the limited capacities of these regulators, especially considering the rapid advancements in AI. The US has not yet implemented a comprehensive federal framework to govern AI through mandatory rules. Instead, it relies heavily on existing legal frameworks, particularly tort law and product liability, to address AI-related harms. Negligence is a key concept in the US approach, holding AI developers liable if they fail to exercise reasonable care in the design, development, and deployment of their AI models, leading to foreseeable harm. The US approach also acknowledges the potential application of product liability to AI models, particularly under the risk-utility test, which assesses whether a product's design poses unreasonable risks in comparison to its benefits. There is no consensus yet on whether AI models should be considered products under US law, which makes the application of product liability less clear-cut.

Any reflection cannot fail to start from a snapshot of the whole system that is evolving, both technically and normatively, at the level of Western legal systems. It is what in the jargon of any risky activity can also be called the Vantasner Danger Meridian, a kind of line that fixes the timeline of the real riskiness. This may determine an alarmist technophobia or bring closer the more desirable pragmatic attitude of the regulator. We understood that AI/R, managed by Frontier AI, imposes on us a (transnational/national) regulation that cannot fail to follow the entire life cycle of the intelligent machine. From the initial development phase, up to testing and then introduction into the market, with periodic checks on the outcomes, whether positive or negative, including in terms of harm to the workers. The more advanced the artificial intelligence, as in the case of Frontier AI, the more risks there are. And this becomes even more true for forms of AI/R that cooperate, interact, coexist with workers. Frontier AI risks may arise from three factors that regulation, including labor law regulation, must somehow intercept[48]. The first factor relates to unexpected problems, generally arising from the fact that Frontier AI presents the actual ability to self-determine. It may do this precisely at the stage of distribution in the marketplace, or in performing tasks for and with end users. The second factor concerns the issue of security in the deployment phase. End-users could demand further development from Frontier AI that is not aimed at legitimate ends, in the sense of legally unjustifiable, which are not entirely foreseeable (see above) and could multiply its harmful scope. The third factor relates to the problem of proliferation, since Frontier AI is generally open-sourced, allowing anyone to

[48] Here we follow the approach that was developed by the group of scholars and experts, whose work became the subject of reflection and a point of reference for the policies of law by the intergovernmental conference dealing with *Frontier AI* in London in 2023, by the European Union, and by the United States of America. See M. ANDERLJUNG, J. BARNHART, A. KORINEK, J. LEUNG, C. O'KEEFE, J. WHITTLESTONE, S. AVIN, M. BRUNDAGE, J. BULLOCK, D. CASS-BEGGS, B. CHANG, T. COLLINS, T. FIST, G. HADFIELD, A. HAYES, L. HO, S. HOOKER, E. HORVITZ, N. KOLT, J. SCHUETT, Y. SHAVIT, D. SIDDARTH, R. TRAGER, K. WOLF, *Frontier AI Regulation*, cit.

develop additional models, even those with purposes that are not legally justifiable. This often involves model theft or cyber-attacks that allow access to patterns that have some strategic value. The diagram that may clarify these factors is taken from the documentation of the AI Safety Summit in London, 2023:



To meet these challenges regulation, including labor law regulation, would be called upon to set (i) international safety standards[49], (ii) enforceable models of compliance with these standards, (iii) transparency and information indices and schemes, and (iii) a macro-regional, no longer just domestic/national, insurance model against Frontier AI-derived labor risk. Let's start with international standards, which should be the object of a form of conventional or treaty regulation, signed at least among the most relevant Western legal systems, including certainly the United States of America, the EU institutions, and Great Britain. Standards should be the subject of study, experimentation, risk auditing, data collection and academic, social, etc. comparison. Auditing schemes aimed at verifying standards should be introduced. This would make it possible to construct standards that are elaborated through the technique of risk assessment, with the weighting of what is actual capacity for danger arising from AI/R (also in the form of the Frontier AI), even at firm level, and its controllability. Assessment by specialized third parties, auditors or the like, would in fact create greater credibility based on what is made known externally, even with the help of such specialized third parties. Hence it would move the construction of standardized protocols on how AI/R should be distributed and what safeguards should be introduced as well as the processes for periodic review of risks and measures to be taken.

But all this would probably not be sufficient in any case. Compliance models should be made enforceable, balancing their impact on potential technological-

[49] In the United States of America, the National Institute for Standards and Technology has created the AI Risk Management Framework. The National Telecommunication and Information Agency has started a path to introduce AI-related insurance policies. In Great Britain, the AI Standards Hub has been established. The European Union has asked the agencies CEN and CENELEC to develop standardized safety models on artificial intelligence.

economic development. This is the real challenge for a lawmaker. An avenue of voluntary adherence to these standards should be promoted, also at the transnational level, with the creation of codes of conduct (self-regulation), certifications, issuance of special prior qualification/license in order to be able to create and then deploy, also for person/machine collaborations, AI/R and Frontier AI systems, supervisory authority actually active in inspections, etc. This would make it possible to activate cross-checking systems in high-tech production units, new occupational safety procedures linked to certified vocational training for each worker collaborating with AI/R, special patents for such workers, etc. Hence it could be possible to build up a compulsory, at least macro-regional, European insurance system, referable to what in Italy is INAIL, for AI/R risks at workplace level. We know that the greater the scope, ordinal and geographic, of an insurance/pension/social security scheme, the greater the benefits for all. The imponderability of this approach specifically relates to the impact on the rule that prevents the harm and risk to the worker as well as allows compensation for the harm under advanced AI/R systems such as Frontier AI. It is an imponderability that stems from the fact that the most appropriate combination of prevention at workplace level and compensation for further harm, in the event of the exercise of employer powers by Frontier AI or any form of artificial intelligence even more advanced, is not yet well known. This determines an impact, also based on interdisciplinary and transnational scientific studies that reference reality, not just the mere idea, such as the one we intend to carry out. With reference to the fact that one cannot fail to verify the adequacy of social-insurance regimes for accidents at work and occupational diseases with respect to the new generation artificial intelligence, adequacy to be measured with respect to the obligation to contribute benefits and the employer's liability regime, while still verifying the efficiency of safety at work systems. These safety at work systems are all or almost all set in the past, not in the future, that is, for a machine subjected (with enormous variability) to the indications of the worker, not for an intelligent machine, Frontier AI, that coordinates, imposes, controls, directs even the worker's performance.

All this represents the most important challenge for the scientific community, the policymaker and, ultimately, for those who work in the industrial relations systems, unions and employer organizations. It is a direction to sense and then decide to follow, knowing that direction is always more important than speed because it represents the whole over the part.

Abstract

Il saggio muove da una constatazione della realtà: la trasformazione tecnologica, dovuta all'intelligenza artificiale di nuova generazione (qui anche "Frontier AI" o "GPAI"), ci costringe a fare una nuova mappatura degli effettivi rischi e delle possibilità di innovazione che derivano dall'interazione tra lavoratori e macchina intelligente. Il che ci pone di fronte a domande più complesse di quelle che ci siamo posti sino a oggi. Qui, a tal fine, si sceglie l'angolo visuale dei nuovi rischi sociali e psico-fisici derivanti da tale interazione persona/macchina intelligente, nella comparazione tra Unione europea e Stati Uniti d'America. Pur essendo convinti che da quella interazione possano derivare certamente maggiore produttività e maggiore benessere, non si può non indagare ciò che sta per accadere nei luoghi di lavoro ri-plasmati dalla tecnologia avanzata, nei quali la nozione classica di rischio probabilmente non riesce più a essere adeguata in ragione della presenza operativa di un terzo elemento tra datore di lavoro e lavoratore. Tale terzo elemento esercita poteri, si confronta con obblighi nonché può determinare danni, creando forme di responsabilità contrattuale e extracontrattuale. Le domande da cui muove il saggio sono le seguenti: cosa desideriamo che l'intelligenza artificiale, anche nella forma di robotica avanzata, nei luoghi di lavoro, faccia per noi, o, meglio, "con" noi, e certamente mai a nostro danno? Chi regola questa interazione tra la persona del lavoratore e l'intelligenza artificiale/robot (qui anche "AI/R")? E come si deve regolare quell'interazione? Inoltre, data tale interazione, cosa potrebbe accadere qualora la responsabilità dell'eventuale danno alla persona del lavoratore dipendesse direttamente/esclusivamente dall'AI/R? L'assicurazione obbligatoria contro gli infortuni sul lavoro e le tecnopatie presenta già oggi una fisionomia per poter comprendere nel nostro sistema previdenziale anche quel danno da lavoro dipendente direttamente da AI/R? E quanto è efficiente, in termini di mitigazione del danno, la procedimentalizzazione che l'attuale regolazione sulla sicurezza sul lavoro impone, anche nell'ambito di produzioni a tecnologia avanzata o dove il lavoratore co-opera con l'AI/R? Il saggio intende avviare il confronto accademico per poter creare un substrato teorico di una nuova branca della disciplina giuslavoristica che attiene allo studio della regolazione dell'intelligenza artificiale/robot in interazione attiva/biunivoca con lavoratori in contesti produttivi tecnologici avanzati (qui definita anche Robot Labor Law – "RLL").

The technological revolution is forcing us to reimagine how we assess the risks and opportunities of human-robot interactions (IWRs) in the workplace. However, the crux of the issue lies not in the technology itself, but in how we, as humans, will integrate AI and robots into our lives and workplaces. We must acknowledge that significant changes are on the horizon, including the emergence of new social and psychological risks associated with human-robot interactions and the most advanced forms of AI (Frontier AI or GPAI). My thesis is grounded in the belief that AI-powered robots (AI/R) should not merely augment human capabilities but should work collaboratively with humans to enhance both productivity and well-being. The absence of a unified scientific framework hinders our understanding of the complex interplay between humans and AI-powered robots in shared physical and social spaces. My research aims to contribute to the development of a new field of labor law, "Robot Labor Law" – RLL, that will address the legal and regulatory challenges posed by human-robot collaboration. To achieve this goal, I will explore the following key questions: What are the desired outcomes of AI and advanced robotics in the workplace? How can we optimize human-AI collaboration? How can we prevent AI/R systems from acting in ways that harm workers? Who should regulate the interaction between workers and AI/R systems? What regulatory mechanisms are appropriate? In the event of a worker injury directly attributable to AI/R, what are the potential legal implications? Do existing insurance systems adequately cover AI/R-related workplace accidents and occupational diseases? How effective are current workplace safety regulations in addressing the unique challenges posed by advanced technologies and human-AI collaboration? Who bears responsibility: the employer, the AI developer, or the AI itself? As AI technologies rapidly integrate into workplaces, how can we ensure the safety of workers interacting with increasingly autonomous AI/R systems? By conducting a comparative analysis of EU and U.S. labor law, I will develop a regulatory framework that safeguards worker well-being in the age of AI-powered robots.

Parole chiave

Diritto del lavoro e relazioni industriali, Robot, Intelligenza Artificiale, Frontier AI, Rischi, Responsabilità, Danni, Risarcimento danni, Danni, Lesioni, Infortuni sul lavoro, Malattie professionali, Misure di mitigazione, Sistemi di sicurezza sociale

Keywords

Labor Law and Industrial Relations, Robot, Artificial Intelligence, Frontier AI, Risks, Liabilities Damages Compensation, Harms, Injuries, Accidents at Work, Occupational Diseases, Mitigation Measures and Insurance Plans, Social Security Schemes