European Journal of Law and Technology, Vol. 15 No. 3 (2024)

EJLT European Journal of Law and Technology

Risk, Harm and Damage as Preset Rational Categories in AI Literature: Do We See or Think the Problem?

Cristina Cocito, Thomas Marquenie and Paul De Hert[*]

Abstract

This article reflects on dominant concepts used in contemporary legal discourse to understand, identify and address problems raised by AI systems, particularly in the GDPR and European Artificial Intelligence Act, that rely on concepts such as risk, harm and damage. Our study questions how far these dominant concepts sufficiently capture the problems AI presents, and whether they guarantee a comprehensive approach to identifying those problems. Building on pragmatist methodologies of problem inquiry (Dewey and Bergson), we argue that while some existing conceptual paradigms may be more suitable than others, they all are located too far ahead in the problem inquiry process, as defined by pragmatists. Existing paradigms for problem-identification, anchored to preset categories of problems, risk marginalising (other) elements, such as feelings, concerns or other problematic issues. This study eventually calls for further research to explore more critically how concepts such as risk, harm and damage are used in literature to map AI systems' problems. This gives rise to a broader call for research to identify methodologies that can pragmatically frame the challenges of AI systems in order to better and comprehensively address the problems they raise today.

Keywords: artificial intelligence, pragmatism, inquiry, risk, harm and damage, human rights

[*] Cristina Cocito and Paul De Hert are affiliated with the Vrije Universiteit Brussel (VUB); Thomas Marquenie is affiliated with KU Leuven.

1. Introduction

> 'It is a familiar and significant saying that a problem well put is half-solved.'
> J Dewey (1938) in *Logic: the Theory of Inquiry*

> 'Until there is language to describe an experience, that experience is not conscious for our culture which (…) is profoundly logocentric, privileging the word, (…) and repressing anything that does not fit into our current language games.'
> G Maxwell (2013) in 'Intellect and Intuition in Henry Bergson'

> 'We shouldn't regulate AI until we see some meaningful harm (…) There has to be a little bit of harm, so that we see what is the real problem.'
> Michael Schwarz, Microsoft Chief Economist to WEF Growth Summit 23

Inquiries into problems, including their identification and resolution, concern every domain of life, at every stage of human development. In the digital domain, inquiries about problems raised by artificial intelligence (AI) systems are as prominent as ever.[1]

The integration of AI in different sectors has changed traditional human environments and social settings, influencing communications, knowledge acquisition, problem-solving, risk management and decision-making. Optimism about its benefits has pervaded much of society.[2] Increasingly, however, feelings have crept in that something with such ubiquitous, large-scale and unrestrained use in numerous domains, particularly in relation to more sophisticated AI systems, is not seamless.[3] Various ethical and legal implications concerning AI's intrusiveness, design, accuracy, reliability, transparency and fairness in processing vast amounts of (potentially sensitive) data, and its role in informing important decisions, have been identified.

Both in literature and policymaking, these (negative) implications have been extensively conceptualised and framed in terms of 'risks', 'harms' or 'damages'. These notions, referred to as 'conceptual paradigms' or 'conceptual lenses' in this paper, serve as a mainstream reference for the identification of problems of AI systems, thereby also shaping how such problems are addressed, influencing strategies and policies implemented to mitigate or solve them. In other words, these concepts are used to diagnose the problematic nature or challenges of AI and – having been integrated into European regulation on new AI developments in the General Data Protection Regulation (GDPR) and the European Artificial Intelligence Act (AI Act) – to regulate AI systems.[4] European data protection legislation addresses data processing

[1] The authors would like to thank Andrés Chomczyk Penedo (VUB) for his valuable feedback on this paper.

[2] Roberto Viola, 'Artificial Intelligence, Real Benefits' (European Commission 2018).

[3] Mass surveillance scandals like Cambridge Analytica raised this awareness, leading to EU data protection reforms.

[4] In this study, we recognise a double role of risk, harm and damage. Firstly, risk, harm and damage are concepts. By conceptual lens or paradigms, we also broadly convey the idea of systems of thought and approaches, used in literature and policymaking, that use such concepts to interpret and give meaning

implications as risks to rights and freedoms resulting from data processing leading to damages.[5] The AI Act extends central regulatory focus to risks and harms to the health and safety and fundamental rights of people.[6]

In this paper, we discuss these conceptual (and epistemic) paradigms used to interpret and regulate problems within AI systems.[7] Our research question asks how far concepts of risk, harm and damage are suitable for identifying and giving meaning to problems raised by AI in realistic and comprehensive terms, and thus to effectively regulate them.[8] Specifically, this paper examines how these concepts are applied within two key legislative instruments: the GDPR and the AI Act.[9] To do so, we turn to philosophical tradition, in particular to pragmatic approaches to problem-inquiry or identification (John Dewey and Henry Bergson). These approaches offer relevant insights into the actions of interpreting or identifying problems that are unknown to individuals based on their experience and practice.[10]

This study highlights more dogmatic shortcomings in the use of risk, harm and damage, leading to a call for more research on their use in literature to identify AI's problems. The status of risk, harm and damage as the most appropriate conceptual lenses in AI's phenomenology is questioned.[11] We reason that how problems are

to problems within AI systems. These concepts shape how people understand and interpret a particular subject and provide a way of analysing problems, guiding the questions and methods used in inquiries. Secondly, these notions are used in legal frameworks to regulate technologies.

[5] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC [2016] OJ L119/1. Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L 119/89 (Law Enforcement Directive) adopts the same approach.

[6] Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L2024/1689.

[7] 'Epistemic paradigms' refers here to frameworks of knowledge and understanding that guide how we perceive, study and interpret reality. Given the broad and ambiguous nature of the term 'technology', this paper focuses on 'AI systems' to provide a clearer scope for the analysis and discussion, hence using 'AI' or 'AI systems' interchangeably to refer to technologies more generally.

[8] An explanation of realistic and comprehensive knowledge is given later in this paper.

[9] While the choice of focusing on the AI Act is straightforward with regards to AI systems, the GDPR impacts AI by regulating personal data collection, processing and AI-driven automated decision-making.

[10] From this main analysis, this study also briefly paves the way for further legal research by presenting two best practices provided in literature. While this study's primary focus is assessing whether these notions fully capture AI's challenges, this final reflection aims to connect the study to future research and prepare the authors for it.

[11] The term 'phenomenology' is used here in a tongue-in-cheek way to refer to how authors mainly write about AI's impacts and people's lived experience with AI, and to elaborate upon more pragmatic methodologies for problem-identification, i.e. methodologies that resonate with practice and complexity of reality. Our phenomenology study, including this paper, refers to a study of prominent methodologies used in literature to identify problems with AI.

diagnosed and recognised in the digital sphere is key for the 'fitness'[12] of (problem-solving) solutions, and thus for the regulation and application of technologies.[13]

This paper is structured as follows. After presenting the pragmatist framework that anticipates our conceptual discussion (Section 2), we define key technologies and terms used in contemporary technology debates, such as AI, algorithms, big data, automated decision-making systems and profiling (Section 3). This allows us to put forward preliminary considerations about the level of granularity when assessing AI systems, based on considerations such as whether some types of technologies are more problematic than others (Section 4). Then we briefly assess the analytical scope of concept of risk (Section 5), of harm (Section 6) and of damage (Section 7) and their use in the GDPR and AI Act. The use of this trinity of familiar legal notions diverges from the approach to problems and problem-identification embraced by pragmatism (Section 8). Thus, in this section we also discuss further practical flaws of these conceptual lenses. To demonstrate the feasibility of better assessment methods in line with pragmatism, we end by providing two best practices found in legal literature that provisionally illustrate more open, flexible and pragmatic methodologies for assessing the negative implications of AI systems: the 2021 UN Report on Privacy in the Digital Age; and the legal analysis of AI by Bart van der Sloot and Sascha van Schendel (Section 9).

2. Pragmatism in Problem Identification

To assess how far the concepts of risk, harm and damage can give a comprehensive account of the problems within AI, we look at the notion of a problem from a pragmatist perspective. In this study, pragmatism offers an alternative method or approach for 'inquirers' to identify these problems – whether scholars discussing and researching the matter in literature, policymakers or AI regulators. It suggests that these issues be interpreted through the real-life experiences of individuals, potentially revealing neglected aspects that are not addressed by traditional concepts or lenses. In this study, it offers a complementary lens to analyse as well as enhance or refine existing approaches or methods.[14]

[12] John Dewey, *Logic: the Theory of Inquiry* (New York Henry Holt 1938) 107.

[13] In this paper we see that risk, harm and damage are often conflated, confusing and obscuring their difference in giving account of the implications of AI systems. When examined closely, these notions reflect different aspects or dimensions of a problematic situation and emerge at different stages of problem identification. From a pragmatist perspective, their use seemingly 'rushes' inquiry and they go too far in problem identification, bypassing other factors that may relate to ethics, moral issues, vulnerabilities, capabilities or feelings, namely concerns or mistrust in AI systems (Section 8). This risks overlooking the richness of AI's implications, limiting processes of understanding reality. Moreover, while theoretically abstract notions open to interpretation without context, they are often given fixed categorisations and 'boundaries' in the digital world that limit their interpretation and regulatory application. In digital-related legal discussions and frameworks, risks and harms are too tied to the idea of human rights risks and harms, which narrows the discussion to a limited set of concerns.

[14] Cristina Cocito, Paul De Hert and Thomas Marquenie, 'Do human rights frameworks identify AI's problems? The limits of a burgeoning methodology for AI problem assessment' (2024) submitted to *International Journal of Human Rights*.

2.1 A Pragmatic Conceptualisation of Problems

According to American pragmatist philosopher John Dewey, a problem originates from a situation that is felt as concerning or unsettling and impairs agents' normal experience or habits within the environment.[15] The inquirer's initial condition simply consists of feeling something wrong, and this could have endless explanations. For Dewey, this situation is somehow 'precognitive',[16] although a necessary condition of cognitive inquiries, and (objectively) 'indeterminate in relation to its future issue'.[17] Thus, it may not necessarily exist within the researcher's conceptual background.[18] Only once the situation is subjected to inquiry can its problematic character become clear and be properly defined. Hence, feelings and guesses by inquirers, as well as imagination, play an important role in appraising problems.[19] The notion of problems is flexible and open, including to new experience, developments, redefinitions and categorisations. As we see in the next section, problems must firstly be felt, guessed and understood from experience and observation.

2.2 Identifying (New) Problems: Methodologies Proposed by Pragmatism

In *Logic* (1938), Dewey distinguished five phases shared by processes of 'inquiry', where inquiry encompasses processes of problem-solving and learning.[20] Problem-identification, for instance, constitutes only one phase of problem inquiry. Before identifying a problem, inquirers (AI researchers, scholars or regulators) first perceive that things are off-balance and 'do not hang together'[21] in their environment. This could relate, for instance, to AI regulators or researchers working on AI governance receiving complaints or noticing unusual hiring patterns that seem unfair or inconsistent before formally identifying specific AI bias in hiring algorithms. At this stage, the exact issue cannot yet be pinpointed, but there is a general sense that something is not working right. This feeling concerns an indeterminate situation of disturbance which prompts 'inquirers' into questioning and knowing further.[22] This is the antecedent of a problem being recognised.

[15] Dewey (n 12) 101–119, 105; see Dewey 's example of hunger, 27.

[16] ibid 107.

[17] ibid 106; DM Mackay, 'What Does Mr. Dewey Mean by an "Indeterminate Situation"?' (1942) 39(6) *The Journal of Philosophy* 141, 145.

[18] Inquiry is indeed aimed at acquiring new knowledge, seen by pragmatism as open-ended, dynamic and receptive to new experiences. Dewey (n 12) 107; Michael Luntley, 'What's the Problem with Dewey?' (2016) 8 *European Journal of Pragmatism and American Philosophy* 1, 8.

[19] Dewey (n 12) 105–108.

[20] Inquiry as problem-solving transforms an indeterminate situation into a determinate one. Dewey identifies five phases: 1) the indeterminate situation as starting point; 2) identifying and locating the problem; 3) defining a problem-solution as an initial idea; 4) assessing the solution's functional fitness; 5) integrating facts and meaning. In this text, only the first three steps are illustrated for relevance purposes. ibid.

[21] ibid 105.

[22] ibid 107; Shane J Ralston, 'What Can John Dewey Teach Us About Everyday Problem Solving?' (2020) Apeiron Blog, https://www.academia.edu/43906103/What_Can_John_Dewey_Teach_Us_About_Everyday_Problem_Solving accessed 2 December 2024. Dewey sees this situation as objectively indeterminate. Yet, inquiry

As researchers move on to the second phase, the indeterminate situation shifts from a feeling of something wrong or concerning to the concrete identification and conceptualisation of the problem. This is similar to processes of disease diagnosis by medical professionals:[23] they begin with general observations, like a patient feeling sick, and progressively narrow down to a specific diagnosis. Doctors might start with a 'gut feeling' that something is wrong based on vague clues, patient history or their general experience with similar cases, and thus proceed by making implicit assumptions and intuitive thoughts. This intuitive judgment often leads them to pursue further tests or alternative diagnoses.[24]

Therefore, to identify the problem researchers need to dwell on the actual and observable 'terms of the problem',[25] investigating experience through common sense and scientific inquiry.[26] This research needs to be pragmatic and open-minded, allowing for an exploration of the situation based on real-world experience rather than rigid or predefined ideas. Problem identification should be sensitive to the overall quality of a situation ('the problem must be felt before it can be stated'),[27] freed from strict conceptual limitations.[28]

It is problem-solving that is primarily driven by concepts, ideas, rational thinking and solutions rather than feelings and observation. Success in problem-solving is contingent upon how accurately problems are understood, so inquirers should not rush problem-identification and give it adequate attention.[29] Only by progressing in inquiry do experiences, suggestions and ideas begin to work in a mutually reinforcing 'cycle'. Yet, sensations and observation should initially institute and feed development of concepts, ideas and categorisations, not *vice versa*.[30]

Henry Bergson, a French philosopher known by Dewey, shared with him a pragmatist and open epistemology based on experience.[31] In *The Creative Mind* (1946), Bergson

has also a partial aesthetic dimension since it is based on inherently felt situations. Larry A Hickman, 'Inquiry: a Core Concept in John Dewey's Philosophy' (1997) 17 *Free Inquiry* 21.

[23] Ralston (n 22).

[24] Similarly, an AI researcher may start having a suspicion that an AI model is disproportionately affecting a certain population group by having an intuitive judgment from interacting with people affected. This researcher may collect data to see whether there is bias and understand the form of bias. Subsequently, he may identify this bias as a (risk to) human rights violation.

[25] Dewey (n 12) 109.

[26] ibid 60.

[27] ibid 70.

[28] Inquiry aims at promoting intellectual growth (the expansion of horizons and formation of new purposes) requiring an attitude of openness and 'intellectual hospitality' for alien or new perspectives, fostering a continued capacity for growth. John Dewey, *Democracy and Education* (The Pennsylvania State University 2001) 182.

[29] Dewey (n 12) 108–109; also Hickman (n 22).

[30] 'It is possible to have the work of observation so controlled by a conceptual framework fixed in advance that the very things which are genuinely decisive in the problem in hand and its solution are completely overlooked.' In this case, 'everything is forced into the predetermined conceptual and theoretical scheme' Dewey (n 12) 70.

[31] Dewey praised Bergson's philosophy while not fully supporting intuition as a way of knowledge. Gerard Deledalle, 'Dewey J. Un inédit de John Dewey: Spencer et Bergson (1965) 70 *Revue de*

distinguished two methods of understanding and knowing reality – analysis and intuition.[32] Analysis, as a way of gaining knowledge, applies preset perspectives and familiar concepts to reality. Since reality is defined in terms of 'elements already known' (human-made artificial concepts),[33] analysis only leads to relative and artificial knowledge.[34] Conversely, intuition, which resonates with Dewey's idea on feelings and observation, does not rely on artificial perspectives or concepts. It is, rather, inclined to acquire disinterested and comprehensive knowledge of things.[35] Because it merges intelligence and instinct, intuition depends on open-mindedness and creativity necessary to perceive how others feel and understand experience.[36]

Both Dewey and Bergson emphasise identifying problems through observation and experience, rather than relying on abstractions or predetermined concepts.[37] Concepts should guide, not constrain, problem identification as these limit our understanding of reality and its complexity.[38]

Pragmatism, with its focus on flexibility, practice and context, is relevant for effective problem-solving in daily life.[39] This approach, and problem identification more generally, is applicable across all areas, including the evaluation of AI's implications – from minors to more troubling ones. Before exploring how these ideas connect with the dominant notions used in AI's phenomenology, the technologies involved by this discussion are introduced.

Métaphysique et de Morale 326; Henry Bergson, *Matter and Memory* (Zone Books 1988) 184–185. Bergson praised pragmatism in Henry Bergson, *The Creative Mind* (Philosophical Library 1946) 248.

[32] Bergson, *The Creative Mind* (n 31) 187.

[33] ibid 190.

[34] ibid.

[35] Intuition means 'intellectual sympathy' towards experience, a conscious activity unaffected from predefined perspectives on things, enabling deeper understanding of experience. Bergson (n 31) 187. Maxwell sees it as 'instinct become intelligent', a phenomenological empiricism beyond words. Grant Maxwell, 'Intellect and Intuition in Henry Bergson' (2013) https://grantmaxwellphilosophy.wordpress.com/tag/affect/ accessed 2 December 2024.

[36] Bergson's knowledge-approach is open, ever-changing and creative. Henry Bergson, *Creative Evolution* (1907) (Cosimo Classics 2005).

[37] Pragmatism, as opposed to constructivism, is not rejecting any positivist notion of objectivity but can be situated in between positivism-constructivism and objectivity–subjectivity extremes at the X and Y axes. The first step in pragmatism is to understand expressions and language – i.e. what do we mean by concepts? – which researchers then take further. 'The core conclusion is that the result of pragmatic research depends on the opening thought of the researcher.' Kuldip Neupane, 'Understanding Pragmatism for Research: Which Pragmatism?' In Rhituraj Saikia and others (eds) *Thinkers: Creating New Ideas of Research* (Eudozia Research Center 2023) 562.

[38] Against this pragmatist background, we can further clarify the idea of knowledge in this paper. Comprehensive knowledge means an understanding that captures the complexity and richness of AI's negative implications (as broad as possible), while objective knowledge is grounded in real-world experience rather than abstractions and theories.

[39] Cocito, De Hert and Marquenie (n 14).

3. Phenomenology of AI-related Technologies: AI, Algorithms, ADM, Profiling and Big Data

Before moving to the core discussion of this paper, we here provide a notional outline of key technological systems relevant to the dominant concepts for AI problem identification. This helps to define the scope or limits of their functioning and application.

Firstly*,* there is artificial intelligence (AI), a loosely defined term frequently used as a conceptual umbrella to encompass various related ideas and technologies. Given its widespread use in different contexts and fields of research, there exists no authoritative or agreed definition of AI. In technical terms, it generally refers to the capability of a piece of software or system to mimic the reasoning, problem-solving and decision-making power of the human intellect.[40] In a more legal sense, the AI Act defines an AI system as a machine-based system designed to operate with varying levels of autonomy and which may exhibit adaptiveness after deployment, and which, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments.[41]

Secondly, there are algorithms, which play such a critical role in developing AI systems. Algorithms refer to pre-instructed operations executed by a computer system to reach a specified outcome or solve a problem drawing from other inputs.[42] Of particular importance in contemporary AI applications are machine learning (ML) algorithms, as opposed to expert-based or knowledge-based systems.[43] Knowledge-based systems rely on trained knowledge data and software to reproduce human judgement and decision-making to achieve certain objectives, generally complementing human decision-making. ML algorithms are more sophisticated AI systems relying on data and models that enable the software to learn and improve from experience and produce more refined outcomes without human intervention. An algorithm can be thus hand-coded by a programmer or generated automatically from data as in ML.[44] ML algorithms increasingly rely on big data – our *third* technology – generally defined as the 'practice of combining huge volumes of diversely sourced information and analysing them, using more sophisticated algorithms to inform decisions'.[45]

[40] European Parliamentary Research Service (EPRS), 'Understanding Algorithmic Decision-Making: Opportunities and Challenges' (5 March 2019) 4. See also Jan De Bruyne and Cedric Vanleenhove, *Artificial Intelligence and the Law* (1st edn, Intersentia 2021) 2.

[41] AI Act, art 3(1).

[42] Thomas H Cormen and others, *Introduction to Algorithms* (3rd edn, The MIT Press 2009).

[43] Expert-based systems are classic knowledge-based systems, so the terms are often used interchangeably.

[44] EPRS (n 40) 4.

[45] EDPS, 'Meeting the challenges of Big Data: a call for transparency, user control, data protection by design and accountability' Opinion 7/2015 (19 November 2015) 7. For another big data definition see Bart Van Der Sloot, 'Is the Human Rights Framework Still Fit for the Big Data Era? A Discussion of the ECtHR's Case Law on Privacy Violations Arising from Surveillance Activities' in Serge Gutwirth, Rene

Algorithms are key to the design and development of our fourth technology – automated decision-making systems (ADM). ADM involve decisions made by automated means, including via algorithms, without meaningful[46] human involvement.[47] An ADM system is essentially a socio-technical framework that comprises a model of decision-making translated into computable code by an algorithm, the data used as an input to feed the system, either to learn from it or analyse it on the basis of that model, and the environment surrounding its use.[48]

Fifthly, there is profiling, a technique frequently used to support decision-making by providing algorithm-derived profiles about individuals and populations. Profiling relies on inferential analysis, often through big data mining techniques (operations within large datasets) and algorithms, which identifies statistical correlations or patterns within datasets used as indicators to classify subjects as members of a cluster.[49] The GDPR only covers profiling based on personal data,[50] but the inferential analysis technique increasingly relies also on proxies and metadata. There are different profiling techniques, varying according to the subjects of profiling and the way these are profiled.[51]

4. Role of Context v Stationary Visions of Problems with AI Systems

Grasping with precision the technical differences among systems is not always easy for non-computer scientists. Structured demarcations are increasingly blurred since these systems are often used in combination and share similar characteristics, such as large-scale data collection, opacity, complexity, autonomy and unpredictability. In fact, some conceptual overlaps between them persist in literature, such as between AI systems, algorithms and ADM or big data and profiling. Because of that operational convergence, legal, ethical and other implications are likely to align.[52] Delineating their

Leenes and Paul De Hert (eds), *Data Protection on the Move: current developments in ICT and privacy/data protection* (Springer 2016) 411, 414.

[46] Guillermo Lazcoz and Paul De Hert, 'Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-Requisites against Abdicating Responsibilities' (2023) 50 *Computer Law and Security Review* 1; GDPR, art 22(1).

[47] Information Commissioner's Office (ICO), Guide to the General Data Protection Regulation (Report) (2018) 148.

[48] Mathias Spielkamp, *Automating Society.* Taking Stock of Automated Decision-Making in the EU (Report) (1st edn, BertelsmannStiftung Studies 2019) 1, 9.

[49] Inferential analysis predicts and categorises behaviour by profiling individuals based on unchangeable (e.g., age) or changeable (e.g., habits) characteristics. FRA, *Preventing Unlawful Profiling today and in the future: A guide* (Handbook) (Luxembourg 2018) 3.

[50] GDPR, art 22(1).

[51] Profiling can be individual (identifying suspects) or group-based (predictive analytics). Group profiling includes distributive and non-distributive types. AH Vedder, 'Het einde van de individualiteit? Datamining, groepsprofilering en de vermeerdering van brute pech en dom geluk' (1998) 3 *Privacy & Informatie*, 115–120, in Gloria Gonzalez and Paul De Hert, 'Understanding the legal provisions that allow processing and profiling of personal data—an analysis of GDPR provisions and principles' (2019) *ERA Forum*, 610.

[52] Katerina Demetzou, 'GDPR and the concept of risk' in Eleni Kosta and others (eds), *Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data* (Springer International Publishing 2019) 137, 149. Therefore, authors often advocate that regulatory responses should also

operational boundaries (i.e. the limits or scope within which different systems function) is however important to avoid disregarding the nature and unique features of all these systems, and to correctly interpret, understand and identify negative implications that may result from their use.

Whereas some systems share similar problematics, other – increasingly sophisticated – systems may be significantly more problematic than others. Within a legal perspective, ML's complexity and autonomy make it more challenging than simple expert-based systems. Because it establishes probabilistic correlations among data rather than causal associations, inferential profiling also renders ADM more challenging in a legal perspective. The application of the same technology may also cause more severe problems in certain sensitive domains, such as healthcare, than in others, such as law enforcement. Further, these may be more problematic within certain socio-political settings, such as countries lacking the guarantees of the rule of law. This illustrates that understanding and identification of AI's problems should be contextual and specific while avoiding too abstract or general assumptions, also in accordance with pragmatist approaches highlighted above.[53]

Furthermore, novel AI developments may generate new types of problems or intensify existing ones.[54] AI's implications are contingent upon and evolve in parallel with the evolution and sophistication of AI. Likewise, AI systems could change in response to post-deployment adaptiveness.[55] This highlights the need to avoid a stationary perspective of AI problems and, conversely, the need for an open, flexible and non-static scope of our understanding of the impacts of AI.[56] In other words, this understanding should be as broad, open-minded or receptive, and dynamic as possible.[57]

As implied by pragmatic approaches to problem identification, this understanding or identification, which influences how we carry out and respond to negative assessments of AI, is likely to be informed by the theoretical lenses we rely upon to identify problems. This includes how we provide meaning to notions of risk, harm and damage in relation to AI. As a preliminary caveat, conceptual paradigms or lenses should be applied with flexibility, without determining problem-identification, so that such implications can be understood objectively and comprehensively.[58] The next sections explore how notions of risk, harm and damage are traditionally used as conceptual

converge. Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Transparent, Explainable, and Accountable AI for Robotics' (2017) 6(2) *Science Robotics*, 3.

[53] Dewey's inductive methodology implies that the specific is studied to draw conclusions of more general applicability. Bergson sees abstractions as limiting absolute knowledge.

[54] Commission, Commission Staff Working Document: Liability for Emerging Digital Technologies, SWD(2018) 137 final, [4.2].

[55] AI Act, art 3(1).

[56] Also Demetzou (n 52) 15-6 on risk as 'non-static' and dynamic concept.

[57] ibid.

[58] Since AI is not only integrated in the environment but also influences how people interface with it, this evidently requires the pragmatic engagement and evaluation of AI's impact based on people experience (Dewey).

paradigms or parameters for identifying problems within AI systems. To do so, we define them briefly in the context of the GDPR and AI Act.[59]

5. Risk as a Problem Yardstick in GDPR and AI Act: Key Insights and Implications

Risk is a central concept in contemporary debates and the regulation of digital and AI systems, used extensively to indicate or capture the problematic nature of AI in most AI 'phenomenology'. It is an abstract rather than a concrete notion that slips out of authoritative definitions.[60] Although being subject of many interpretations across disciplines,[61] risk can be neutrally defined as 'a combination of the probability [likelihood] of occurrence of a defined hazard and the magnitude [severity] of the consequences of the occurrence'.[62] Working Party 29 (WP29) – the predecessor to the current European Data Protection Board – speaks of risk in terms of 'a scenario describing an event and its consequences, estimated in terms of severity and likelihood'.[63]

The GDPR implements a risk-based approach to the regulation of data processing activities.[64] Although GDPR's provisions do not explicitly define risk, its definition relies on WP29's statement on risk.[65] The harmful dimension of risk in the GDPR, thus the scope of GDPR's risk, is determined by violation of rights and freedoms of natural persons.[66] Risks thus relate to (potential) negative implications for people's rights,

[59] The authors wish to emphasise that the following sections are necessarily brief and not exhaustive, acknowledging that fully unpacking each of them would require a dedicated paper on its own.
[60] By 'abstract' we mean that it is not something that can be easily quantified or understood in a simple, direct way without being applied in specific contexts, and is open to interpretation.
[61] Ortwin Renn, *Concepts of Risk: a Classification* (Universitat Stuttgart 1992).
[62] Fredrick Warner, *Risk: Analysis, Perception and Management: A report of a Royal Society Study Group* (The Royal Society 1992) 4; Raphael Gellert, *The risk-based approach to data protection regulation* (OUP 2020) 27.
[63] Art 29 WP, 'Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679' (2017) WP 248, 15. In regulatory governance on technology, two main meanings or uses have been made of the concept of risk: as a harmful scenario (in Jeroen Van Der Heijden, 'Risk as an approach to regulatory governance' (SAGE 2021) https://journals.sagepub.com/doi/full/10.1177/21582440211032202 accessed 2 December 2024, 5; Milda Macenaite 'The "Riskification" of European Data Protection Law through a two-fold Shift' (2017) 8 *European Journal of Risk Regulation* 506, 508)) and as a method for assessing and managing harmful events (Van Der Heijden, 5). For a typology of risk in regulatory governance see Julia Black, 'The role of risk in regulatory processes' in Robert Baldwin, Martin Cave and Martin Lodge (eds), *The Oxford Handbook of Regulation* (OUP 2010).
[64] On 'risk' in the former EU Data Protection Directive see Macenaite (n 63) 8.
[65] WP29 (n 63); WP29, 'Statement on the role of a risk-based approach in data protection legal frameworks' (2014) WP 218. By pinpointing the three GDPR risk elements (event, harm and risk factors), Gellert notes that the event element is often overlooked in risk assessments. Raphael Gellert, 'Understanding the notion of risk in the General Data Protection Regulation' (2018) 34 *Computer Law and Security Review* 279, 281. On GDPR's risk see also Karen Yeung and Lee A Bygrave, 'Demystifying the modernized European data protection regime: Cross-disciplinary insights from legal and regulatory governance scholarship' (2021) *Regulation & Governance* 137.
[66] GDPR, art 35.

particularly their right to data protection.[67] The problematic dimension of risk in the GDPR is dictated by the quality of data processing, and thus it is connected and contingent upon the (damaging) event element of risk.[68]

The AI Act also adopts a proportionate, risk-based regulatory model but combines it with a precautionary approach to AI. Its definition of risk aligns with its common interpretations as the 'combination of the probability of an occurrence of harm and the severity of that harm',[69] yet focuses on harm instead of a hazardous or damaging event. Authors note that, unlike the GDPR, the AI Act's risk-based approach places greater focus on, or is more directly intertwined with, the harm element of risk.[70] Risks to fundamental rights (harm) are the regulatory yardsticks or benchmarks determining which AI situation is more problematic and should be forbidden.[71] Too risky AI applications are therefore prohibited in the AI Act, and requirements and obligations for AI providers depend on the level of risk posed to health, safety and fundamental rights. While the initial consultation on the AI Act proposed a sector- and case-specific approach,[72] the main reference or scope of the risk in AI remains that to fundamental rights harm.

6. Harm for AI Problem-Identification in GDPR and AI Act: how is it Applied?

Although it essentially consists of different conceptual components, the notion of risk is frequently applied in an elastic manner, in particular to give expression to its various elements including the harmful consequences.[73] For some authors, risk can 'mean different things to people, especially in subjective domains like privacy'.[74] This often results in risk and harm being (erratically) conflated to represent a unique problem to be tackled. Nonetheless, even while risk typically carries a threat-like connotation related to possible undesired or hostile events, the event having *ex ante* risks may, or may not necessarily, end up having harmful consequences.

[67] WP29 (n 65) 4.

[68] Therefore, so do risk assessments (WP29 (n 65)). On the role of risk as risk compliance emphasising the quality of processing and the occurrence of (damaging) events caused by GDPR non-compliant activities, see Gellert (n 65).

[69] AI Act, art 3.

[70] Raphael Gellert, 'The role of the risk-based approach in the General data protection Regulation and in the European Commission's proposed Artificial Intelligence Act: Business as usual?' (2021) 3(2) *Journal of Ethics and Legal Technology* 15, 24–28.

[71] AI Act, art 7. This distinguishes between four levels of risk upon the severity of harm on people: unacceptable risk; high-level risk; limited risk; and minimal or no risk.

[72] Explanatory Memorandum to AI Act, 3.

[73] Among its three elements, risk is said to be less often conveyed as indicating the underlying situations or events. Gellert (n 65).

[74] CIPL, 'Risk, High Risk, Risk Assessments and Data Protection Impact Assessments under the GDPR' (2016) (Report) CIPL GDPR Interpretation and Implementation Project, 13. CIPLS notes that, for instance, a data breach may be a risk, but it does not always harm data subjects, which depends on data usage. Alternatively, risks like loss of confidentiality or financial loss are resulting harms or damages from a data breach.

Like the notion of risk, harm is also a vague and broad term. In *Harm to Others* (1987), Feinberg defines harm as 'the thwarting, setting back, or defeating of an interest'.[75] Harm also generally defines an adverse or upsetting outcome of an event or situation which generally embeds experienced impairment, suffering or loss, and which can take numerous forms.[76] Feinberg's definition also includes harm as damage (an extended, derivative type of harm) and harm as wronging in a normative sense.[77]

The GDPR does not refer to harm but to risks to the rights and freedoms of natural persons, i.e. fundamental rights violations, as well as material and immaterial damage.[78] It seemingly overlooks the issue of harm and is also more focused on the event linked to data processing, and less or not at all on its consequences (harm).[79] Otherwise, the GDPR could also appear to conflate the notion of risky processing and harm within the conceptual structure of 'risk', such that risk and harm are essentially the same matter.[80]

The AI Act explicitly focuses on harm or adverse impact, particularly on health, safety and fundamental rights, as the key factor in regulating AI systems.[81] Harm to fundamental rights provides the crucial condition under which an uncertain AI situation is determined as problematic, a risk, and thus forbidden. AI's problematic nature is thus defined and scaled upon the likelihood (risk) of fundamental rights harm. In fact, Article 6 states that 'an AI system (…) shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons'.

While the initial proposal for the AI Act gave a more narrow account of harm caused by AI, the final version more broadly refers to material or immaterial harms, and specifically mentions harm of physical, psychological, societal or economic nature.[82] As noted above, in both the GDPR and the AI Act the terms of reference for identifying problems with AI systems remain primarily linked to violations, adverse effects, or harm to fundamental rights and freedoms.

[75] Joel Feinberg, *The Moral Limits of the Criminal Law Volume 1: Harm to Others* (OUP 1987), 3; for a definition of societal harm see Nathalie Smuha, 'Beyond the Individual: Governing AI's societal harm' (2021) 10(3) *Internet Policy Review*. For public order crimes see Larry J Siegel, *Public Order Crimes in Criminology* (7th edn, Wadsworth Publishing 2000).

[76] Physical harm (injuries), psychological harm (mental abuse), material harm (damage), immaterial harm, legal harm, social harm and economic harm.

[77] Although harm as interest-thwarting and normative harm are more often combined, e.g. legal or fundamental rights harm, not all harms as invasion of interest are wrong. Feinberg notes that consensual harm, such as self-inflicted harm, is excusable. Feinberg (n 75) 5.

[78] 'The risk-based approach in the GDPR goes beyond a narrow "harm-based-approach" that concentrates only on damage.' Rather, it considers every potential and adverse effect, from individual impacts to broader societal consequences. WP29 (n 65) 4.

[79] Gellert (n 70) 22.

[80] CIPL (n 74).

[81] AI Act, Recital 5, 16, 27, art 7.

[82] AI Act, Recital 5. Also Recital 155 and 33 for other types of harm relevant in the AI Act.

7. The Role of Damage for Problem-Identification in GDPR and AI Act

Damage is another concept used to capture negative implications associated with data-driven and AI systems. While damage and harm are often treated as synonyms,[83] the two do not necessarily always have the same meaning. For instance, damage is also relatable to harm to things 'when they are objects of no one's interests'.[84] In its strict sense, damage mainly entails tangible impairment to inanimate objects rather than people, except in cases of economic or reputational damage.[85] In legal terms, damage describes losses, injuries or harms stemming from wrongful or negligent actions to property, individual or reputation.[86] While legal damage means a legal right violation, fundamental rights harm is instead rarely defined in terms of damage.

The GDPR links risks to rights and freedoms to damages caused by data processing.[87] The GDPR covers physical, material and non-material damage resulting from infringement of the regulation.[88] Whereas the text does not provide a definition of non-material damage, it seemingly interprets the term in a broad way.[89] Conversely, the AI Act does not explicitly refer to damage, but the proposed AI Liability Directive provides the legal framework about damage and harm arising from the use of AI systems.[90] While the proposed Directive does not define damage, damage and harm seem likely conflated in this instrument.[91]

[83] Damage can be considered harm when someone has an interest in the damaged object or in its normal functioning. Here, damage is harm in its derivative meaning. Feinberg (n 75) 3.

[84] ibid.

[85] While it can be both material and unmaterial, damage refers to physical 'harm caused to something which makes it less attractive, useful or valuable'. *Oxford Dictionary*, https://www.oxfordlearnersdictionaries.com/definition/english/damage_1.

[86] Damage may not always coincide with legal (right) harm (*damnum absque injuria*), and legal harm (*injuria sine damnum*) can occur without physical damage. William S.C. Goldstein, 'Standing, Legal Injury Without Harm, and the Public/Private Divide, (2017) 92(5) *NYUL Review* 1572. In criminal law, endangerment offences involving wrongful conduct are punishable without them leading to tangible harms.

[87] GDPR, Recital 75.

[88] GDPR, art 82, Recital 75.

[89] GDPR, Recital 146. The GDPR definition of damage includes deprivation of rights or control over personal data (Recitals 75, 85). Damage is linked to fundamental rights violations, like discrimination. EU case law (Case C-300/21, *UI v Österreichische Post AG*) also limits non-material damage to genuine harm, not mere upset derived from violations of data protection law.

[90] Commission, Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM(2022) 496 final. The Directive covers damage caused by AI outputs or failures (Recital 15). See also Explanatory Memorandum on the AI Liability Directive Proposal for the relationship between the AI Act and AI Liability Directive on managing risks and damages (p 2).

[91] 'Liability provisions apply where AI Act's provisions on preventing risks (of harm) to fundamental rights have failed, thus providing compensation also for fundamental rights harm.' ibid p 9.

8. A Critical Pragmatist Assessment of Risk, Harm and Damage

8.1 A Pragmatist Assessment of Risk: Overtaking Problem-Identification Steps

Now let us look at the three dominant concepts in light of our pragmatist essentials (Section 2). Our critical assessment considers these notions from two angles: first, as stand-alone conceptual lenses; and second, in relation to fundamental rights.

We start with the notion of risk. Based on the discussion above, it is manifest that the notion of risk does not only conceptually differ from harm and damage, but also resonates poorly with the vision of a problem by pragmatists. Moreover, risk as a lens for problem-identification aligns poorly with pragmatist approaches to problem-identification and their guidelines. Why?

'Risk' refers to a scenario implying a possibility (the risk) that a defined event (often called hazard) with certain consequences (harm or damage) will occur.[92] Risk is thus *ex ante* harm's possibility, meaning it deals with possibilities or uncertainties before any actual harm occurs, while harm and damage are *ex post* facts, which means they refer to real outcomes that have already happened. Despite often being used interchangeably, or strictly together, the three notions refer to different dimensions or aspects of a (problematic) situation that are relevant and can be recognised at different moments, thus implying different dimensions of problem identification. In practice, they also call for different types of assessments and regulatory responses, with risk being usually prevented with anticipatory measures while harm and damage require remedial action. When related to our pragmatist essentials, the concept of risk weakly resonates with pragmatist understanding of problems.

Firstly*,* risk does not relate to something concretely happening, but something that has a chance of happening in the future. This means that risk is always a hypothetical condition, involving possibilities or scenarios that have not yet materialised and may not do so.[93] This hypothetical character of risk is at odds with Dewey's problem, which is rooted in lived experience, tangible disruptions and concrete situations that can be felt and observed. While risk is speculative, a pragmatist idea of a problem emphasises disruptions that directly impact real-world situations. As a lens to look at problems, risk is not well suited for such experiential investigation, as it deals with potential rather than actual and experiential scenarios. Hence, it does not entirely fit pragmatic approaches to problem-identification defined in Section 2.

Secondly*,* although risk implies degrees and elements of indeterminacy in regard to future events, the scenarios implied or described by risk are defined and determinate.[94] In other words, the event and outcome implied by risk must be known to AI researchers to various degrees, thus be specific and determined. This enables risk to be converted

[92] Renn (n 61) 56.

[93] Gabriela Argüello and others, 'Introduction to Regulation of Risk' in Abhinayan B. Bal and others (eds) *Regulation of Risk. Transport, Trade, and Environment in Perspective* (Brill Nijhoff 2022) 6.

[94] Conversely, while risk involves partial indeterminacy, a problematic situation is fully indeterminate at first and cannot be understood without inquiry (Dewey, n 12).

into quantifiable probabilities, be measured, and foreseen with relative certainty. Likewise, this permits risk assessments to be turned into practical measures and solutions to handle or mitigate the occurrence of events and their consequences.

In reiterative terms, risk by definition presupposes a determination of a problem because it is formed both by a hazardous event and its undesirable consequences (otherwise a risk would not exist). Risk's identification implies that certain problems (harm or damage contained by its definition) are already identified and formulated, to which risk refers. Therefore, it is preliminarily essential to identify the unwanted event and the consequences it may generate – the problems – before any risk identification (and assessment) can be pursued.[95] The notion and use of risk as a conceptual lens for identifying AI's problems thus place inquirers beyond the initial stages of inquiry (namely feelings of indeterminacy and process of turning indeterminacy into a concrete problem through experience investigation).

8.2 A Pragmatist Assessment of Harm and Damage (and Risk): Rushing Inquiry

Let us now consider the notions of harm and damage. These notions embed actual events that with different levels of severity impair individuals' actions or practices. The manifestations of both harm and damage are often observable and can be objectively assessed. As conceptual paradigms for AI's negative implications, harm and damage align better with Dewey's vision of problems as (tangible) disruptions in experience. However, a pragmatic vision of problem also seems conceptually broader and more inclusive than harm (e.g. to human rights), as well as damage, although these concepts may be part of that understanding. In other words, harm and damage can be seen as strands of a broader 'problem conceptual repository'.

Dewey's 'problem repository' is open-ended and creative.[96] Thus, problems cannot be minimised only to forms of harm and damage. As conceptual paradigms, both harm and damage have established features discussed above (Sections 6 and 7) that, despite being broad, define and close the boundaries of these concepts. Harm and damage are thus sufficiently generous in representing a vast array of problematic situations linked to AI – and EU law generally interprets them broadly – but they are seemingly not sufficient in themselves as paradigms to capture all issues. They may be more appropriate than risk, but identification cannot be confined to them. The open nature of problems, and of researching problems, in pragmatism suggests that there may be other sentiments, concerns, disturbing issues or events that do not easily fit the notion of harm and damage within the meaning discussed above.[97] These considerations are

[95] Though this paper does not focus on 'hazard', problem identification, or the original problematic situation, may more naturally revolve around the hazardous event which refers to an actual change in circumstances possibly causing harm (Gellert n 73) (this change also aligns with Dewey's view of problems as change in experience). See Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation No 1907/2006, OJ L 353/1. Hazard makes risk inherently problematic due to its potential for harm.

[96] Bergson's rich vision of reality and knowledge reflects this.

[97] Consider the example of self-driving cars, where forms of more intuitive concerns may be present. Though statistically safer than human drivers, people may still feel apprehensive or uncomfortable

also linked to the human rights argument that will be illustrated in the following section.

Through a pragmatist lens, too much focus on concepts like risk, harm and damage may oversimplify or 'rush' problem identification and processes of understanding complex issues, bypassing other factors that may relate to ethics, moral issues, vulnerabilities, capabilities or other feelings, namely concerns or mistrust in AI systems.[98] While a problem may exist without (causing) harm or damage, they are always outcomes of a problem or problematic event. In other words, harm and damage can be manifestation of a problem as negative effects or outcomes. Thus, these or other issues anticipate the identification of harm or damage but risk remaining overlooked if too much focus is placed on harm and damage.[99]

From these arguments it follows that risk, harm and damage as dominant conceptual lenses for representing AI's problems may push inquirers too far into later stages of inquiry, namely in the second and third stage of inquiry where problems are already defined and identifying solutions relies on concepts and reasoned or systematic categorisation. Overemphasising these lenses as a dominant methodology may overlook other aspects in pragmatism – feelings, perceptions, concerns and open-ended experience investigation[100] – narrowing the scope of what is considered in a problematic situation. This narrowing has regulatory implications, as it may lead to a failure to address the full spectrum of issues – particularly emotional, psychological or social effects – that could not fit neatly into traditional notions of risk, harm or damage. We review this argument in the following section.

about relinquishing control to an AI system. This intuitive and felt discomfort might stem from the fear of technology failure or malfunctioning, ethical dilemmas (e.g., how the car will manage unavoidable accidents), or loss of control. If regulations only focus on measurable risks (such as accident rates) or consider harm as human rights violations, this deeper intuitive mistrust of AI may be overlooked. Work automation is also illustrative. Beyond the financial impact of job loss, workers fired by AI may experience feelings of loss of purpose, dignity or identity. These emotional and psychological feelings, though significant, may not fit into traditional notions of harm like financial damage or rights violations. For instance, industrial workers replaced by AI may not only face economic hardship but also feel a profound loss of community and meaning in their lives – an aspect not always addressed in discussions of economic or human rights harm.

[98] ibid. See Section 10 for an example of the relationship between fundamental rights and ethics and the idea of vulnerabilities.

[99] To give another basic example, the opacity of certain technologies' operations does not fit the notion of harm nor damage (nor risk). Likewise, it does not necessarily entail suffering, loss or injury. However, it is a problematic issue, insofar as it deviates from transparency standards, or a damaging condition of data processing that may harm (violate) people rights. In this sense, the approach in the GDPR better resonates with this position as it emphasises fixing the quality of the processing rather than addressing harm.

[100] This involves examining factors causing disturbance; for Dewey, the inquirer's role is to make sense of this initial unsettling components in experience.

8.3 Using Risk and Harm Combined with Human Rights Narrows the Analysis

Risk and harm – although broad concepts in principle – have not only conceptual limitations in conveying all possible problems raised by AI systems, but also more practical ones due to their almost exclusive relation to fundamental rights in EU law.

Due to the wide acceptance of human rights as moral benchmarks and a problem-solving framework in Europe,[101] most issues related to AI systems are linked to and considered through their lenses.[102] Therefore, problems tend to be seen as pointing to specific rights at times, meaning that risks and harms are primarily interpreted in terms of potential rights violations.

This human rights perspective has been the subject of other publications by these authors and exceeds the scope of this study.[103] It suffices to say, however, that human rights recognise issues, wrongs or harms within the confines established by human rights norms.[104] Despite their evolving interpretation, these norms are generally regarded as finite, meaning that a human rights-based representation of problems is both inclusive and exclusionary.[105]

[101] Hin-Yan Liu, 'AI Challenges and the Inadequacy of Human Rights Protections'(2021) 40 *Criminal Justice Ethics*; Cocito, De Hert and Marquenie (n 14).

[102] Cocito, De Hert and Marquenie (n 14).

[103] ibid.

[104] For this argument see also Liu (n 101).

[105] Despite the evolving interpretation of human rights by courts such as the ECtHR, human rights in international instruments are largely deemed as finite by human rights scholars (see Hurst Hannum, *Rescuing Human Rights: A Radically Moderate Approach* (CUP 2019)). Concerns about expanding human rights can be challenged with the evolving nature of EU fundamental rights as peculiar to certain jurisdictions (v. human rights as norms of global character) or across scholars of different schools of thought (e.g. human rights as natural law v. human reason). Human rights have indeed accommodated evolving needs and interests over time, as testified by environmental considerations, the EU Charter right to data protection or the CJEU-created right to be forgotten. However, new rights creation is often a gradual process shaped by political compromises rather than swift or radical developments. Celeste notes that new rights development, especially in the digital realm, is often uneven and incomplete, with 'anemic' constitutions struggling to encompass new phenomena (see Edoardo Celeste, *Digital Constitutionalism: The Role of Internet Bill of Rights* (Routledge 2023) 212). Expanding human rights is constrained by legal principles, historical precedents and societal consensus. Legal codification is also slow, while new rights require legitimacy relying on collective consent (for these arguments see Cristina Cocito and Paul De Hert, 'Relying on Digital Principles to Complement Existing Rights. A Human Rights Assessment of the 2022 European Declaration on Digital Rights And Principles', in Ben Wagner and others (eds), *Research Handbook on Human Rights and Digital Technology* (2nd edn, forthcoming)). This means that these frameworks are generally rigid and do not easily accommodate the novel or complex challenges of AI: the pace of adapting to new challenges reflects careful consideration and negotiation. Additionally, not every concern or value can be translated into a human or fundamental right, but there must be limits to how far rights can expand to address evolving issues. Also, human rights frameworks are limited in addressing broader structural issues like poverty or global challenges (Cocito, Marquenie and De Hert, n 16). Some scholars propose alternatives, such as Nussbaum's focus on human capabilities, to address broader concerns beyond traditional rights (Martha C. Nussbaum, 'Capabilities as Fundamental Entitlements: Sen and Social Justice' (2003) 9(2–3) *Feminist Economics*. Simultaneously, proposals for new online rights, such as a right to disconnect, may testify that, for various reasons, human rights are not able to cover everything. For an explanation of this see Cristina Cocito and Paul De

By referring back to a (pre)determined problem framework (and also a limited one), namely human rights law, the notion of risk and harm as interpreted by EU law is too fixed and narrow as a problem-identification methodology in light of pragmatism. This warns that fixed frameworks marginalise elements needed to understand problems, restricting both identification and resolution. The relatively rigid nature of rights, which translates into a fixed or specific scope of risks and harms when related to human rights, contrasts with the necessity for flexible, dynamic frameworks that can capture evolving problems of advancing technology in changing contexts. In this sense, the conceptual structure of damage, despite often being conflated with harm, may seem more flexible and freed from fixed rational categories set out in advance in EU law, since it is not generally tied to fundamental rights.

This 'closed' and predetermined approach to problem-identification more generally collides with pragmatist precepts based on open, 'disinterested' and creative perspectives in researching problems. Put differently, this focus is not only too conceptual but also narrow and leaves little room for other concerns and new forms of problems.[106]

Social scientists often underline that what is perceived as negative, a danger or a threat, and hence the way individuals provide concrete meaning to notions of risk, harm or damage is often shaped by values, influences and socio-cultural contexts.[107] However, these conceptual lenses may imply abstract visions that do not fully align with the objectivism and open-ended nature of pragmatic inquiry based on real-experience investigation. As such, human rights may only offer limited instruments for problem-identification.

Although having a more practical dimension when used as an assessment method, authors also highlight the conceptual and practical difficulties in combining the notion of risk as an assessment method to that of fundamental rights, and thus on the incompatibility of risk-based approaches with fundamental rights. Risk-based approaches are in fact only one among many possible approaches to technology.[108] That argument is primarily linked to the idea of measurability, which collides with fundamental rights. According to Yeung and Bygrave, human rights violations cannot be quantified and scaled because of their higher moral value grounded in individual

Hert, 'The transformative nature of the EU Declaration on Digital Rights and Principles: Replacing the old paradigm (normative equivalency of rights)' (2023) 50 *Computer Law and Security Review*.

[106] See pragmatism's intellectual hospitality (Section 2). This 'predetermined' nature means that fundamental rights are defined in legal texts and national constitutions. They are often interpreted within the confines of the original text and legal precedents, limiting flexibility in addressing new issues, e.g. those raised by AI. Judges base their rulings on established legal frameworks and past jurisprudence, adapting to contemporary challenges but within the boundaries of these predefined norms.

[107] Mary Douglas, *Risk and Blame: Essays in Cultural Theory* (Routledge 1992) 40.

[108] See OECD's principle-based approach and the CoE's right-based approach to AI. Sometimes, the EU uses a market-based approach to AI.

dignity.[109] While violations may vary in severity, human rights are binary, meaning that something is legal or illegal; it is a human rights violation or it is not.

Conversely, measurability is an inherent element of risk, and risk assessments generally rely on mathematical and statistical estimates of the probability and severity of quantifiable or tangible harm.[110] Rhetoric about risks may suggest that fundamental rights violations can be measured and quantified.[111] Human rights can yet be restricted only in narrow circumstances according to certain requirements (e.g. necessary and proportionate in a democratic society) that must be evaluated in judicial review, not within risk assessments and not by data controllers possibly ill-trained to do so.[112] Furthermore, risk-based regulatory approaches assume that risks cannot be fully eliminated.[113] However, the notion of permitting harmful consequences not entirely erasable on fundamental rights does not fit human rights doctrine, thus remaining dogmatically impure.[114] Likewise, it is difficult to scale human rights protection, meaning that it collides with cost–benefit analyses typical of risk-based approaches.[115]

Risk is not only about statistical assessment, but also implies cost–benefit, trade-off evaluations with a utilitarian aim to maximise benefits and minimise costs. In AI, this often entails balancing the potential risks for individuals (e.g. loss of privacy, bias in decision-making) against overall advantages within decision-making processes. An AI developer may thus proceed with the development and deployment of AI systems that offer to bring advantages such as improved efficiency despite a slight risk for data or privacy. Risk assessment and cost–benefit analysis thus often allow certain acceptable harm if the perceived advantages outweigh them. Human rights law does not permit such trade-offs. Human rights are deemed inalienable, and violating individuals' rights cannot be justified by providing benefits to others. This is rooted in the idea that certain rights, like dignity, privacy, equality and non-discrimination, are non-negotiable. Under human rights law, even one case of discrimination can be unacceptable, requiring accountability and oversight.

Additionally, it is often argued that using human rights harm as a paradigm for AI's problem-identification is inadequate to capture all possible types of the negative implications of AI. Human rights harm traditionally presumes that there is an identifiable violator causing harm, and an identifiable victim.[116] In the context of AI systems operating on vast scale, in opaque manners, or without human intervention, this presumption becomes increasingly obsolete.[117] The individual nature of harm

[109] Yeung and Bygrave (n 65) 10.

[110] ibid.

[111] ibid.

[112] ibid.

[113] Macenaite (n 63) 512.

[114] This is a different discourse than proportionality of rights' limitations.

[115] Macenaite (n 63) 521.

[116] Tania Krupiy and Jaqueline McLeod Rogers, 'Mapping Artificial Intelligence and Human Intersections: Why We Need New Perspectives on Harm and Governance in Human Rights' in Aoife O'Donoghue and others (eds), *Research Handbook on Global Governance* (Edward Elgar 2023); Cocito, De Hert and Marquenie (n 13).

[117] ibid.

linked to fundamental rights, as such envisaged by the AI Act, has also notably become untenable in relation to AI mass-scale operations. For example, AI can generate widespread, systemic effects that harm multiple individuals or communities.[118]

It is not within the scope of the present paper to assess these arguments in detail. Rather, they remain open for further research and discussion. Moreover, they do not mean to discard the importance of these concepts in contemporary AI debates. They served to illustrate their limitations – as the mostly used lenses in AI's phenomenology – in comprehensively and realistically giving account of AI's problems, while leaving open the prospects for more pragmatic and open methodologies for AI problem-identification. These are, in our view, possible and within everybody's reach. In the following section, we briefly present two approaches to AI's problems taken from literature that provide an embryonic illustration of better, more pragmatic practices of problem-identification.

9. Best Practices Guiding Further Research

9.1. UN High Commissioner, Problematic Issues, Harm to Rights and Risks (Best Practice 1)

In the 2021 Report, The Right to Privacy in the Digital Age, the UN High Commissioner assessed the implications of AI systems used by states and businesses, considering practice and various contexts of application.[119] The report firstly identifies concerning features or issues of AI based on experience, including data exploitation, large-scale identification, probabilistic inferences, intensified vulnerability, inaccuracy and opacity. It then illustrates how these concerns 'are experienced in practice' across different sectors of AI application (criminal justice, public services, employment and online content management).

By linking those issues to specific AI applications, the report identifies relevant risks for human rights while considering other problems not directly associated to human rights risks, harm or damage. In law enforcement, AI predictive tools associated with the probabilistic nature of predictions may flow (risk) into harm for privacy, fair trial, freedom from arbitrary arrest and detention, and the right to life. Meanwhile, issues around opacity in AI raise pressing questions in terms of accountability in law enforcement. In employment, data exploitation in AI monitoring creates risks of privacy harm. However, other AI issues may lead to function creep and lack of accountability.

This approach illustrates a multifaceted type of problem identification. It illustrates how various of AI's concerning features or issues are firstly identified and examined in practice with respect to different AI applications (in line with our discussion in Section 5). Relying on human rights to identify problems with AI, these issues lead to

[118] Smuha (n 75) 23; AI Act, art 7(2)(d) only refers to the severity of harm, including AI's capacity to affect a plurality of persons.

[119] UNHCR, The Right to Privacy in the Digital Age. A/HRC/48/31 (2021); Cristina Cocito and Paul De Hert, 'United Nations: AI can pose risks to human rights' (2021) 197 *Privacy Laws & Business*.

the report identifying relevant risks and harm for human rights. However, the report does not examine, or frame, impacts associated with AI systems within the confines of particular rights. Moreover, this human rights focus does not prevent the UN from collectively and more openly assessing other relevant issues raised by AI, considering concerning conditions but also other problems that do not fit either the notions of (human rights) harm or risk. These concern, for instance, practical and technical dimensions of AI. The UN's approach thus better resonates with pragmatism, which emphasises practice and identifies a broad range of AI implications.

9.2 Van der Sloot and Van Schendel: Substantive Risks and Procedural Issues

In assessing the use of big data in the Dutch public sector, legal academics van der Sloot and van Schendel distinguish between substantive risks and procedural issues.[120] Substantive risks refer not only to human rights harm but also to public values. These risks arise from scenarios associated with use of big data that may impact the exercise of material rights and substantive justice, as well as undermine values. These include citizens limiting and conforming their behaviours due to fear of surveillance (*chilling* effect) or enhanced social inequality through biased big data process (*Matthew* effect).[121]

Along with this substantive dimension of big data problems, these authors identify ten procedural issues related to big data, including the impossibility of individuals knowing about data processing, the weakened legal standing of individuals against big tech, lack of individual damage, and the collective nature of interests affected by big data targeting. These big data procedural issues may impact individuals' ability to enforce rights in a procedural view.[122]

The representation of the implications of big data resonates with our previous discussion by firstly identifying scenarios and technological problems that create risk of harm to human rights, both in a substantive and procedural view.[123] By combining these dimensions, the authors go well beyond a narrow identification of big data issues in terms of human rights risks or harm. These problems relate to, but are not limited to, (human rights) harm and risks, and include issues in terms of public values or issues of a procedural nature. This approach more pragmatically identifies AI's implications and diversifies forms of disruption, issues and risks, including but not limited to human rights harm.

[120] Bart van Der Sloot and Sasha van Schendel, 'Ten adjustments to Dutch procedural law in light of the data-driven society and autonomous systems' ('Tien aanpassingen aan het Nederlands procesrecht in het licht van de datagedreven samenleving en autonome systemen') (2020) *Tijdschrift voor Toezicht, AFl*. 1.

[121] More in van der Sloot and van Schendel (ibid).

[122] ibid.

[123] ibid.

10. Conclusion: 'Harm' and 'Damage' are better than 'Risk', but we still do not See the Problem

Problems with AI systems have been mostly identified through notions of risk, harm and damage. The relevance of these concepts is undeniable, but exclusively relying on them for AI problem-identification limits our view. This familiar triad of legal concepts fails to comprehensively represent problems within AI. There may exist other, better suited lenses to understand AI-problems and, to explore these, our paper draws insights from John Dewey's theory of pragmatism in problem-identification, favouring a naturalistic approach that views knowledge as arising from an active adaptation of the human organism to its environment.

Approaches to understanding problems within AI need to be flexible and non-static. Our lenses to read these problems need to evolve and be capable of accommodating changes, and evolving problems and contexts in which AI systems are applied; and this should reflect experience. When we try to make sense of the challenges posed by AI, we need to avoid using strict or fixed definitions.[124] Notions such as risk, and to a lesser degree damage and harm, were problematised as recommended conceptual lenses for problem-identification. We considered these notions in the GDPR and AI Act, without exhaustively assessing their use in policymaking and literature.

First, there is the notion of risk. In AI regulation, the focus on risk is central but does not always capture the full extent of a problematic AI situation. An appropriate process of problem identification needs to identify the relevant elements at each stage of the process. The primary focus of researching problems should be devoted to the identification of the constitutive elements of risk, namely the hazard and consequences of that event. Then, there is the use of harm and damage. Though closer to the idea of problems, their use does not encompass all negative implications of AI. These cannot be reduced to mere notions of harm and damage; our vision should be much broader than the conceptual horizon embedded by these concepts. Conversely, harm and damage could be a manifestation of problems. Overall, all these notions seem to place AI researchers ahead in inquiry, dismissing a deeper investigation of what initially underpins problem identification, like feelings, concerns, or other more straightforward issues or disturbing facts.

Our analysis of methodologies to identify problems within AI is not final; this article is only part of a broader attempt to establish a phenomenology of AI. Further research is needed on how AI's negative implications are conveyed in literature. A flavour of such a broader endeavour is given by our short presentation of two approaches as best practices guiding future research: the UN Report on Privacy in the Digital Age and the AI work of van der Sloot and van Schendel. These studies are fine, albeit embryonic, illustrations of more pragmatic problem-identification adopting different

[124] At the same time, this flexibility must be balanced with the need for legal certainty. Regulatory frameworks should be clear and predictable to ensure that all parties involved understand their responsibilities while still allowing room for adaptation as AI and its implications develop. Achieving this equilibrium is essential for adaptability and a stable legal environment.

paradigms or lenses, and embedding a more multidimensional perspective on AI's negative implications. These analyses, picked to provide examples, allow the consideration of ethical, technical and legal questions, broadening AI problem identification. Focusing beyond risk, harm and damage may, in fact, also imply the development or improvement of alternative ways whereby AI-related problems can be assessed and identified, potentially incorporating more experiential and collective approaches.

The two selected best practice cases are particularly relevant for current academic literature that sees human rights as a privileged (meta) vantage point for looking at law. Our study demonstrates that correct problem-identification may be impaired by the almost exclusive fundamental rights focus of risk and harm as conceptual lenses (while concerning less the paradigm of damage). This narrow lens can hinder a flexible, perceptive and open-ended inquiry thereby marginalising other elements that may be important to fully understand problems. Given that human rights derive from ethics, appeals to ethics – the initial mainstream approach in AI – may broaden the analysis.[125] Some authors talk instead of vulnerabilities in AI systems as exploitable weaknesses or predispositions to harm.[126] These vulnerabilities or other softer issues can then inform regulatory frameworks. Additionally, we illustrated that the notion of human rights harm is increasingly inadequate to give account to all problems associated with AI operations.

Human rights were and are developed as solutions for past problems, and although new human rights can be defined and older rights can be updated through interpretation by courts, this is time-consuming; we want to, and should, understand today's problems.[127]

[125] Ruxandra Andreea Lapadat, Trusting the Use of AI in the Law Field: Foremost, an Ethical Challenge, in Jan Klučka, Lucia Bakošová and Luboslav Sisák (eds), *Artificial Intelligence from the Perspective of Law and Ethics: Contemporary issues, perspectives, and challenges* (Leges 2021) 45–46.

[126] ibid.

[127] Compare: ' It should be a commonplace, but unfortunately it is not, that no education—or anything else for that matter—is progressive unless it is making progress. Nothing is more reactionary in its consequences than the effort to live according to the ideas, principles, customs, habits or institutions which at some time in the past represented a change for the better but which in the present constitute factors in the problems confronting us. The fact that a given change was made in order to realize a desirable end in view signifies that the life-conditions before and after are different. In the process of attaining that good, a new situation was created. A new complex of life-conditions was brought into existence presenting its own distinctive characteristics and problems. Blind attachment to what was good for a state of affairs that no longer exists prevents recognition of the needs of the present and blots out of view the desirable ends that those needs should generate. As Emerson puts it, the attained good tends to become the enemy of the better. New problems cannot be met intelligently by routine application of ideas and principles which were developed in solving different problems […]' (John Dewey, ''Introduction to The Use of Resources in Education written by Elsie Ripley Clapp' (1952) in Martin S. Dworkin (ed), *Dewey on Education* (Teachers College Press 1959), 95. This passage continues with the reassuring message that the foregoing does not mean a complete rejection of the old – on the contrary.