Juanita Goicovici *

REBUTTABLE PRESUMPTIONS OF CAUSALITY AND REVERBERATIONS OF EVIDENCE DISCLOSURE, AS EPITOMIC PIECES IN THE PHYSIOGNOMY OF LIABILITY FOR DEFECTIVE AI

Abstract

The paper examines the typification of the rebuttable presumptions of causality the use of which has been introduced by the new "Proposal for a Directive of the Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence" ("AILD Proposal") while insisting on the quintessence of the binomial set of presumptions aimed at facilitating the probatory efforts and resorting to a "disclosure of evidence mechanism" in tort litigation founded on AI deficiencies generating bodily harm and/or patrimonial losses to consumers. As resulting from the provisions of Art. 1, para 1 and 2 of the AILD Proposal, the 'hybridisation' of the conceptual nature of civil liability applied in matters concerning autonomous / embedded AI defectiveness does not remain deprived of potential consequences in terms of the reserved possibility of judicial courts, to assess the existence of the illegal action/omission of the responsible persons by referring to compliance with the transparency obligation incumbent on AI providers, as resulting from Art. 13 of the Draft Regulation (EU) on AI (Artificial Intelligence Act) (i) or to retain the civil liability of manufacturers/importers or AI providers for bodily/patrimonial damages caused to consumers, in the assumptions concerning the ignoring, at the time of AI system design or development, of the necessity of "human effective surveillance" for the time bars related to the AI usage (ii).

JEL CLASSIFICATION: K13, K39, K41

SUMMARY

1 Foreword and introductive observations – 2 Binary taxonomies of rebuttable presumptions of causality under tort liability regimen – 2.1 Presumption of causal nexus between defendant's faulty conduct and damageable AI result – 2.1 Presumption of non-observance of the AI provider's duty of care – 3 Presuming causality in liability litigation pillared on damageable AI deficiencies / AI flaws – 3.1 Administration of evidence in liability claims conducted towards the providers of "high-risk" AI systems – 3.2 Recourse to the presumption of causality in claims against AI non-professional users – 3.3 Dichotomic approaches to the administration of evidence in liability cases concerning medium/low-risk AI systems – 4 Courts' ordering on the disclosure of relevant information on AI flaws and deficiencies – 4.1 Proportionality of the judicial measure – 4.2 Temperament (concerning the disclosure of evidence order) pertaining to the consequences on third-party extra-patrimonial or commercial interests – 5 Conceptual and practical interrogations on the

* Associate Professor PhD. at the Faculty of Law of the Babeș-Bolyai University of Cluj-Napoca.

pertinent use of presumptions of causality in cases concerning 'self-learning' / 'self-evolving' AI algorithmic categories – 6 Concluding remarks

1 Foreword and introductive observations

The contemporary efforts to scrutinise the 'classic' liability regimes' capacity to metabolise and absorb civil liability rules adapted to 'harm causing' artificial intelligence, are constantly intensifying, while the physiognomy of specific liability[1] for autonomous or incorporated AI defective products are considerably influenced by their 'regimenting' into 'risk categories' scaled from 'highly threatening' to 'lower degrees' of menacing.

Allogeneic for this type of civil liability remains the presence (and recourse to) rebuttable presumptions to fill evidentiary gaps that could make particularly vulnerable the procedural posture of the consumer injured by the defective autonomous or embedded AI action/omission or by the interaction with algorithmic systems showing design flaws/manufacturing defectiveness. As resulting from the provisions of Art. 1 para 1 and 2 of the AILD Proposal, the "hybridisation" of the conceptual nature of civil liability applied in matters concerning autonomous / embedded AI defectiveness does not remain deprived of potential consequences in terms of the reserved possibility of judicial courts, to assess the existence of the illegal action/omission of the responsible persons by referring to compliance with the transparency obligation incumbent on AI providers[2], as resulting from Art. 13 of the Draft Regulation (EU) on AI (Artificial Intelligence Act) (i) or to retain the civil liability[3] of manufacturers/importers or AI providers for bodily/patrimonial damages caused to consumers[4], in the assumptions concerning the ignoring, at the time of AI system design or development, of the necessity of "human effective surveillance" for the time bars related to the AI usage (ii).

Secondly, the "allopathic" conceptual approach, which appeals to the AILD Proposal by consecrating a "binary" set of presumptions of causation in the perimeter of civil liability for AI harmful deficiencies[5], is rendering permeable the border between 'classical' civil responsibility based on culpable behaviour / gross negligence and, on the other versant of the discussion, objective responsibility[6], completely detached from the element of

[1] Mihailis Diamantis, 'Who Pays for AI Injury?' (Oxford Business Law Blog, 04 May 2020) <https://blogs.law.ox.ac.uk/business-law-blog/blog/2020/05/who-pays-ai-injury> accessed 20 October 2023.
[2] Ajay Agrawal, Joshua Gans and Avi Goldfarb, *Máquinas predictivas. La sencilla economía de la inteligencia artificial* (Barcelona: Editorial Reverté 2019) 42, 48.
[3] María Luisa Atienza Navarro, *Daños causados por inteligencia artificial y responsabilidad civil* (Barcelona: Editorial Atelier 2022) 74, 81.
[4] Jan de Bruyne, Orian Dheu, and Charlotte Ducuing, 'The European Commission's Approach to Extra-Contractual Liability and AI – A First Analysis and Evaluation of the Two Proposals' [2022] SSNR Electronic Journal <https://ssrn.com/abstract=4239792> accessed on 15 July 2023.
[5] Hélène Christodoulou, 'La responsabilité civile extracontractuelle à l'épreuve de l'intelligence artificielle' (Lexbase Hebdo édition privée 2019 n 807. hal-03349668f) <https://hal.science/hal-03349668/document> accessed 12 July 2023.
[6] Jia Qing Yap and Ernest Lim, 'A Legal Framework for Artificial Intelligence Fairness Reporting' (2022) 81 (3) Cambridge Law Journal 610, 644.

culpable behaviour[7], due to the imperative to facilitate the administration of evidence in compensatory actions[8] initiated by the prejudiced consumer[9], whose interests have been harmed by interaction with deficient AI systems. Starting from the findings on the vulnerability of the procedural position of the plaintiff in 'classical' actions, the admitting of which would require proof of the connection relating the incidence of product defectiveness to the prejudicial effects that were generated, disclosure of elements of proof of "high-risk AI systems" seems to enable the claimant to plead for tort liability based on non-contractual fault (i), while placing the burden of proof on the shoulders of professional defendants, in the perimeter of providing access to AI systems whose complexity often exceeds (including) the predictions of its creators (for AI systems from self-learning / self-evolving AI taxonomy). Consistent with the desiderata of 'unburdening' consumers in providing relevant evidence on damage causation while embarking on non-contractual civil law actions (based on fault / culpable behaviour) related to AI system deficiencies[10], consecrating courts' possibilities to order disclosure of evidence on professional defendants[11] (ii) remains the key premise in understanding the innovative system proposed in the AILD Proposal.

Thirdly, instead of focusing on a more permeable or 'fluid' approach to the subjective element of culpable behaviour as a central pillar for retaining the producers' responsibility, or that of the importer or supplier of the deficient AI, the implications of using a presumption relative to the existence of imputable behaviour would allow the consumer, as claimant, to engage in remedial actions without being placed in the undesirable position of not meeting the particularly difficult evidentiary demands regarding AI behaviour as a source of bodily/patrimonial damage; the proposed regulatory norms are permitting professionals the reversing of the presumption of causation in B2C relations, starting from the premise that the relevant information regarding the nexus of (non)causality[12] is rather exclusively in their possession than in the possession of the

[7] Margo Bernelin, 'Intelligence artificielle: une proposition de directive sur la responsabilité civile extracontractuelle' Dalloz actualité (22 Nov. 2022) <https://www.dalloz-actualite.fr/flash/intelligence-artificielle-une-proposition-de-directive-sur-responsabilite-civile-extracontract> accessed 12 July 2023.
[8] Christoph Busch, 'Rethinking Product Liability Rules for Online Marketplaces: A Comparative Perspective' (The 49th Research Conference on Communication, Information and Internet Policy, 22-24 September 2021) <https://ssrn.com/abstract=3897602> accessed on 15 July 2023; Christoph Busch, 'When Product Liability Meets the Platform Economy: A European Perspective on Oberdorf v. Amazon' (2019) 8 (5) Journal of European Consumer and Market Law 173.
[9] Mehmet Fatih Burak, 'Effects of Artificial Intelligence on E-Commerce' in A N Özker (ed), *Reviews in Administrative and Economic Science Methodology, Research and Application* (Livre de Lyon Publishing 2022) 91, 100.
[10] Miriam Buiten, Alexandre de Streel and Martin Peitz, 'The law and economics of AI liability' (2023) 48 Computer Law & Security Review.
[11] Laurène Mazeau, 'Intelligence artificielle et responsabilité civile: Le cas des logiciels d'aide à la décision en matière médicale (2018) 1 Revue pratique de la prospective et de l'innovation 38, 42.
[12] Alan Butler, 'Products Liability, and the Internet of (Insecure) Things: Should Manufacturers Be Liable for Damage Caused by Hacked Devices?' (2017) 50 (4) University of Michigan Journal of Law Reform 913.

profane consumer[13]. The pertinence of causality presumptions is worth discussing, on the binary premises (that remain partly subjective, generated by the element of faulty conduct and partly objective, focused on the element of the presence of the design/ manufacturing defect of the AI system) of engaging in the civil liability of producers/importers and the contrast established by reference to 'classical' versions of subjective responsibility; especially, the discussion may be conducted from the perspective of consecrating a taxonomy of defects covered by the substantial sphere of civil liability, divided into manufacturing defects versus design flaws and informative deficiencies[14]. Deriving from the scope of penalising the non-fulfilment of the transparency obligation incumbent on the manufacturer/importer of the defective AI systems, to which the defects are added in an autonomous manner, there are four types of defective AI response that may reverberate on the civil liability regimen, including the one consisting in jeopardising the safety of the data uploaded by the consumer, as a distinct species of damage whose coverage will be envisaged by the adapted civil liability regime for the remediation of damages caused by defective AI conduct.

2 Binary taxonomies of rebuttable presumptions of causality under tort liability regimen

2.1 Presumption of causal nexus between defendant's faulty conduct and damageable AI result

Saliently, the mechanism described in Article 4 of the AILD Proposal, which provides for a reversible presumption of causal proximity between the defendant's fault or culpable conduct or inexcusable negligence, remains focused on presuming the existence of a causal nexus; the latter's applicability would be differentiated on grounds related to the range of potentially risky behaviour[15], under which the AI systems have been regimented. In the foreground of the discussion, when engaging in the evaluation of elementary premises for retaining the specific liability of the AI manufacturer/importer, the latter's faulty conduct could also be presumed[16] by national courts based on failure to comply with a judgment on judicial disclosure or preservation of evidence pursuant to the provisions of Art. 3 para 5 of the AILD Proposal. Although the presumption of causation

[13] Aileen Nielsen, 'How to measure and regulate the attention costs of consumer technology' (TechStream: Brookings Institute, 4 November 2021) <https://www.brookings.edu/techstream/how-to-measure-and-regulate-the-attention-costs-of-consumer-technology/> accessed 12 July 2023.
[14] Antonia Waltermann, 'On the legal responsibility of artificially intelligent agents. Addressing three misconceptions' (2021) 3 Technology and Regulation 35, 43 <https://techreg.org/article/view/10985/11959> accessed 12 July 2023.
[15] Tiago Sérgio Cabral, 'Robotics and AI in the European Union: opportunities and challenges' (2018) 4 (2) UNIO - EU Law Journal 135, 146.
[16] Tiago Sérgio Cabral, 'Liability and artificial intelligence in the EU: Assessing the adequacy of the current Product Liability Directive' (2020) 27 (5) Maastricht Journal of European and Comparative Law 615, 635.

gains in relevance[17] in hypotheses, in which the inexcusable omission or culpable action of the AI manufacturer/designer has influenced in a decisively obvious manner the poor performance of the AI system[18] (which would be assessed based on the circumstantial peculiarities of the case), the claimant is expected to produce evidence of existence of the damage. Facing the arduous task of bringing consistent evidence that, in consideration of the damageable results generated by the AI system collapse, or the failure of the AI to provide for an appropriate result, the consumer would still be expected to bring proof of the causal link between the registered damageable result and the alleged dysfunctionality of the AI system, in the perimeter of interactions with AI systems regimented in the medium/low-risk categories; reversibly, concerning damages generated by "high-risk AI systems", the text of Art. 4 para 4 of the AILD Proposal establishes an exception to the necessity of proof by the plaintiff consumer of said causal link, establishing the (relative) presumption of causality. Except in the (rare) cases where the professional defendant proves that, in absence of recourse to the presumption of causation, sufficient evidence and an appropriate level of expertise are reasonably accessible to the complaining consumer, the latter is seemingly enabled to prove the existence of the causal link between the faulty behaviour of the AI system and the recording of the physical/property damage.

Controversy is fuelled by the fact that the incidence of the exception from the sphere of incidence of the causality presumption is subject to the courts' decision involved in settling the dispute, which could find it excessively difficult for the claimant to provide evidence as to the causation of the damage even in cases of medium-risk AI use; obviously evidentiary difficulties will be assessed bearing in mind the characteristics of AI technologies, describable as autonomous conduct, or AI opaque behaviour, rendering the consumer's assignment of explaining the peculiarities of an AI system almost impossible to accomplish from the perspective of the availability of evidence, since the claimant is (almost invariably) facing difficulties to prove (in absence of such a reversible presumption) a sufficient nexus connecting the consumer's decision concerning the selected design parameters for the manufactured AI product[19] and the biological / patrimonial harm caused to the consumer in the interaction with the defective AI system[20]. In terms of the proportionality requirements for ordering the disclosure or preservation (by the defendant AI manufacturer / supplier) of evidence regarding the

[17] Ryan Calo, 'Artificial Intelligence Policy: A Primer and Roadmap' (2017) 51 UC Davis Law Review 399, 435.
[18] Tiago Sérgio Cabral, Iakovina Kindylidi, 'Sustainability of AI: the case of provision of information to consumers' (2021) 13 (21) Sustainability <https://doi.org/10.3390/su132112064> accessed 23 October 2023.
[19] Juanita Goicovici, 'Matricea răspunderii civile extracontractuale pentru prejudiciile cauzate de produsele cu defecte de manufacturare, între testul riscuri-beneficii și testul așteptărilor legitime ale consumatorului' [The Matrix of Non-Contractual Civil Liability for Damage Caused by Products with Manufacturing Defects, between the Risk-Benefit Test and the Test of Legitimate Expectations of the Consumer] (2022) 67 (1) Studia Universitatis Babes Bolyai-Iurisprudentia 106, 185.
[20] ENISA (European Union Agency for Cybersecurity), *Cybersecurity of AI and Standardisation*, March 2023, <https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation> accessed 12 July 2023.

compliance / dysfunctionality of high-risk AI systems[21], it is worth mentioning that the courts' assessment of the plausibility of the consumer's complaint remains essential in the scope of the analysis undertaken a priori by the courts, which may order that the defendant be bound to disclose pertinent evidence regarding suspected "high-risk AI systems", even if this evidence incriminates the debtor of the obligation, by way of derogation from the procedural *actor incumbit probatio* principle (given the fact that the lay consumer does not, most often, possesses evidence that is conclusive or relevant regarding the placement of AI system malfunctions at the time its design/manufacturing protocols were selected); the central question of the plausibility of the claim encapsulates the need, for national courts[22], to establish the existence of sufficient grounds for ordering the disclosure or preservation of evidence regarding "high-risk" AI systems the behaviour of which is suspected to be linked to the occurrence of the prejudice[23].

2.2 Presumption of non-observance of the AI provider's duty of care

The launching on the market of defective AI, encompassing AI design defectiveness, as well as manufacturing flaws, might represent per se a form of violating the duty of care incumbent on the AI providers, both at the manufacturing stage, as well as at the pre-contractual stage, in B2B or B2C contextualised relations; the machine-to-machine (M2M) or AI-to-AI contractual interaction might also raise questions on the pertinence of by design and by default selections operated by the AI manufacturer, in terms of consumer safety[24], or through the lens of the 'risks – benefits' balance. Exploiting the *ex-ante* mechanisms of the AI producer's 'duty of care' remains crucial for the success of the liability systems, both at national and transnational levels.

Moreover, in the continuous process of identifying solutions for entailing AI producers' or AI importers' liability for harmful defectiveness, the problems themselves are subject to evolving, thus requiring higher degrees of adaptability for 'classical' remedial paradigms.

The triptych of defect categories englobed in the objective sphere of the specific liability of AI producers/importers can be distinguished on three levels (to which the fourth is added, consisting of 'loss/damage or alteration of consumer's data'): (i) the objective liability, covering AI manufacturing defects, as defects due to human errors,

[21] Michael P. Chatzipanagiotis, 'Product Liability Directive and Software Updates of Automated Vehicles' (Proceedings of SETN 2020 – 11th Hellenic Conference on Artificial Intelligence, 2020) <https://ssrn.com/abstract=3759910> accessed on 15 July 2023.

[22] Juanita Goicovici, 'The inapplicability of personal exceptions between joint debtors and creditors, under Romanian and French private law' in *Looking for New Paths in Comparative and International Law* (Societatea de Stiinte Juridice si Administrative 2021) 85, 98.

[23] Geraint Howells, Christian Twigg-Flesner, 'Interconnectivity and Liability: AI and the Internet of Things' in Larry A DiMatteo, Cristina Poncibò and Michel Cannarsa (eds), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge University Press 2022) 179, 199.

[24] Juanita Goicovici, *Dreptul relațiilor dintre profesioniști și consumatori* (Hamangiu 2022) 183.

generating AI defects that appear or manifest in the 'design – production – distribution' chain (AI structural defects in the design of algorithmic systems, defects or failures of AI equipment etc.); (ii) types of extra-contractual liability for design defects of AI, or for excessive risk compared to the benefits, in which the respective product incorporating AI was designed in manners ignoring the balance of risks and benefits for consumption[25]; (iii) in the absence of deficiencies in the first two categories, extra-contractual liability can be generated by informational deficiencies[26], consisting of incomplete, inadequate or erroneous information regarding the product assuming autonomous or embedded AI, as described in Proposal COM/2021/206 ("Artificial Intelligence Act"). We argue that flawed or erroneous decisions generated by algorithmic systems, as well as low resistance to performance alterations of autonomous/embedded AI in products, can be preferentially treated as product malfunctions, thus allowing the consumer to implicate the manufacturer/importer in damages specific to AI systems (while preserving the duality of 'autonomous material damage vs. derivative material damage', as well as the dichotomous approach of the categories of 'liability for product security deficiencies' versus 'AI design defects').

The dynamics of adapting the 'classic' regime of "civil liability" for physical/bodily damages caused by products presenting design/manufacturing defects to the legal challenges triggered by the gradual generalisation of the use of autonomous/embedded AI are characterised, above all, by aspects such as the adaptation of the burden of proof, by placing some significant sequences of the provision of evidence in the charge of the professional (located, as a rule, in the position of the defendant in the action initiated by the injured consumer) (i) or even by postulating some (relative) presumptions of causality between the existence of the physical/property damage invoked by the claimant and the presence of the defect in the AI product/AI components (ii). In a carefully calibrated dosage, the incidence of the relative presumption of connection is to be confined to the perimeter of cases where, as highlighted in Recital (25) of the preamble of the Draft Directive on the adaptation of the civil liability regime (…), non-compliance with the due diligence obligations incumbent on the manufacturer or on the designer of the AI system, would set premises for retaining producers' responsibility for damage coverage.

[25] Mihailis Diamantis, 'Vicarious Liability for AI' [2021] SSNR Electronic Journal <https://ssrn.com/abstract=3850418> accessed 12 July 2023.
[26] Juanita Goicovici, 'The Traders' Liability for Lack of Conformity of the Digital Content and of the Digital Services, as Regulated by Directive (EU) 2019/770' (2020) 66 (1 Suppl.) Analele Științifice ale Universității Alexandru Ioan Cuza din Iași, seria Științe Juridice 79, 98.

3 Presuming causality in liability litigation pillared on damageable AI deficiencies / AI flaws

3.1 Administration of evidence in liability claims conducted towards the providers of "high-risks" AI systems

At the horizon of the special civil liability whose specific regime is outlined in Article 7 of the AILD Proposal remain the B2C relationships involving the use or interaction of the consumer (in the sense of natural person, using the disputed product in a predominantly extra-professional context) with actions/omissions of algorithmic systems or deficient autonomous/incorporated AI[27], without the need to establish the existence of a contractual link connecting the defendant and the plaintiff in the action brought before the courts; the extra-contractual nature of the liability remains decisive (in view of the fact that, in most cases, the injured consumer did not contract directly with the designer/ manufacturer of the defective AI). Nevertheless, being closer to a tort liability regime, the special liability the regulating of which is envisaged by the AILD Proposal does not completely distance itself from the subjective element of the fault of the (potentially) responsible subjects[28]. Despite the fact that it does not suddenly or dramatically alter the (preponderantly) subjective nature of the pre-existing liability regimes in the member states, at the level of repairing damages caused by the illegal/imputable or inexcusable action/omission, the harmonised set of rules on liability for AI defectiveness 'dilutes' the mentioned subjective nature accordingly, by consecrating a relative presumption of causality applicable against the responsible persons[29], in the presence of malfunctions of the AI systems whose origin can be placed at the time of AI designing / manufacturing of the algorithmic systems suspected to be at the origin of the damage. The absence/presence of a direct or indirect/ successive contractual link between the responsible person and the consumer injured by the action/omission of the AI system is not crucially relevant, since it will suffice to locate the defect at the time the producer decided on the launching of defective design/defective manufacturing of the harmful autonomous AI (including AI from the self-learning category[30])/ detrimental embedded AI, and even for prejudicial generative AI.

[27] Mikołaj Domagała, 'Threats associated with the introduction of autonomous vehicles as an example of the negative effects of the development of artificial intelligence' in Luigi Lai, Marek Świerczyński (eds), *Legal and technical aspects of artificial intelligence* (Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego 2021) 247, 264.
[28] Mark Coeckelbergh, 'Artificial Intelligence: Some ethical issues and regulatory challenges' (2019) 1 Technology and Regulation <https://doi.org/10.26116/techreg.2019.003> accessed 12 July 2023.
[29] Martin Ebers, 'Liability for Artificial Intelligence and EU Consumer', (2021) 12 Journal of Intellectual Property, Information Technology and Electronic Commerce Law 204 <https://www.jipitec.eu/issues/jipitec-12-2-2021/5289/ebers_pdf.pdf> accessed 12 July 2023.
[30] Adrien Bibal, Michael Lognoul, Alexandre de Streel, Benoît Frénay, 'Legal requirements on explainability in machine learning' (2020) 29 (2) Artificial Intelligence and Law 149, 169.

Would resorting to the subjective element of the fault of the producer/supplier of the AI (at the stage of designing/ manufacturing / distributing the defective AI) sabotage the effectiveness of the non-contractual civil liability remedies placed at the disposal of the injured consumers? Would it suffice, as a premise for the engagement of specific liability, to consecrate the objective element, of the existence of the manufacturing/ conception defect (such as design flaws, cybersecurity flaws, transparency flaws etc.) of the AI system suspected to be at the origin of the bodily/property damage that was subject to B2C litigation? The positioning of the liability of producers, suppliers, or importers of AI as types of extra-contractual liability was not doubled, in the text of the AILD Proposal, by eliminating the subjective element of the fault of the (potentially) responsible persons; therefore, using the term 'culpable' behaviour (of the AI producer/supplier) would not be inappropriate to describe the specific extra-contractual engaging of AI producers' liability. Congruently, it has been emphasised in the Preamble of the AILD Proposal that the new set of harmonised rules on liability for defective AI is meant to complement the existing liability (detached from fault) system applicable to producers and importers for damages caused to consumers, or for certain types of biological (physiological) or psychological[31] damages, as well as for patrimonial damages caused to consumers; the AILD Proposal would cover, at the antipode, the issues of compensating damages caused to various types of victims, although not falling under the definition of 'consumer'. These discrepancies in terms of personal sphere of incidence, as well as in terms of objective versus subjective criteria of liable conduct might be seen as complementary, thus offering a plethora of specific mechanisms in view of compensating the damages caused by defective AI.

Recital (3) of the Preamble of the AILD Proposal notes that the reference to the provider's fault (proven by the complaining consumer, whose interests were harmed by defective AI systems) in the light of the requirements applicable in the matter of the administration of evidence in order to engage in retaining the 'classic' civil liability in national legal systems would be a pernicious solution in the context of the permanently amplified technical complexity of AI systems. Thus, one solution would be to abandon the 'traditional' vision of producers' liability that enshrines the *actor incumbit probatio* principle; more pertinently, and strictly for the field of administration of evidence regarding AI defectiveness, it is crucial to consecrate the conveying, in a first stage, of a relative presumption of connectivity (or of the existence of the causal nexus) between the damage caused to the plaintiff consumer and the action/omission of the expected AI responses, the latter not being exempted from the requirements of proving bodily/property damages. Obviously, the current national rules on liability, especially

[31] Chiara Gallese, 'Legal Issues of the Use of Chatbot Apps for Mental Health Support' in *Highlights in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation. The PAAMS Collection* (Springer Cham 2022) 258, 267.

those based on fault, are not adequate for the settlement of actions in liability for damages related to defective AI conduct.

The substantial changes brought to the 'classic' regime of civil liability of producers/importers or suppliers in the retailer category (as a backup solution), in order to adapt its rules to the legal complexity presented by situations where the damage is generated by the defectiveness of the autonomous/incorporated AI can be listed as follows:

(a) the identification of AI product categories using as benchmarks the degree of risks presented for the bodily or physiological safety of consumers or for the respect of their fundamental rights[32] (being targeted: respect for human dignity, combating discrimination[33] in the case of automated decisions which involve the total or partial algorithmising of the decision-making process, or respecting the right to privacy)[34];

(b) imposing the obligation, for AI in the 'high risk' category, to proceed to conducting an *ex-ante* conformity assessment procedure (which remains optional in the case of the other AI categories);

(c) reporting by the courts, in the event of litigation, of the relative legal presumption of causality between the defendant's culpable conduct and the harmful result by reference to the consumer's justified assumptions or the inability of the AI system to generate results on the adequacy of which the consumer could have reasonably relied, as emphasised in Art. 4 of the AILD Proposal; nevertheless, maintaining the possibility for the responsible defendants to bring evidence to the contrary, of the excusable character[35] of the behaviour that is the subject of the litigation, respectively to bring proof of the non-existence of fault for the imputed actions/omissions, referring to the prejudicial dysfunction of the AI systems, remains a major trait of the harmonised liability regime;

(d) establishing the possibility, for the courts, to refer in the assessment of the illegal act of the responsible persons, to the existence/non-existence and to the degree of adequacy of the measures employed by the responsible person (the AI producer or supplier) within the management system of the risks and the results achieved[36], where

[32] Mark Fenwick, Paulius Jurcys, 'From Cyborgs to Quantified Selves: Augmenting Privacy Rights with User-Centric Technology and Design' (2022) 13 (1) Journal of Intellectual Property, Information Technology and Electronic Commerce Law <https://www.jipitec.eu/issues/jipitec-13-1-2022/5512> accessed 12 July 2023.

[33] Florian Teleaba, Sorin Popescu, Marieta Olaru and Diana Pitic, 'Riscurile bias-urilor observabile și neobservabile în inteligența artificială utilizată în predicția alegerii consumatorului' [Risks of Observable and Unobservable Biases in Artificial Intelligence Predicting Consumer Choice] (2021) 23 (56) Amfiteatru Economic 104, 121 <https://www.amfiteatrueconomic.ro/temp/Abstract_2981.pdf> accessed 12 July 2023.

[34] It has become a truistical observation that the EU AI Liability Act is focused on dividing the product categories into 'high-risk AI', 'medium-risk AI', and 'low-risk AI' (initially, in the last category, were regimented the virtual assistants from the range of chatbots that assist consumers in online contracting).

[35] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, Madhulika Srikumar, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI' [2020] Berkman Klein Center for Internet & Society <http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420> accessed 12 July 2023.

[36] Enrico Francesconi, 'The winter, the summer, and the summer dream of artificial intelligence in law' (2020) 30 Artificial Intelligence and Law 147, 161.

the AI system uses techniques that involve training models[37], as mentioned in Art. 10 para 2 to 4 of the Draft Regulation (EU) on Artificial Intelligence;

(e) enshrining the possibility of the courts to assess the existence of an illegal action/omission by the responsible persons by referring to compliance with the transparency obligation incumbent on AI providers, as stated in Art. 13 of the Draft Regulation (EU) on Artificial Intelligence;

(f) consecrating the possibility of the courts to retain the civil liability of AI producers/importers or suppliers for bodily/property damages caused to consumers, and establishing the possibility for the courts to relate, in assessing the relevance of the AI producer/supplier's liability, to the degree to which the AI system was respecting accuracy or pertinency as emphasised in Art. 15 and Art. 16, let. a) of the Draft Regulation (EU) on Artificial Intelligence;

(g) establishing the possibility for the courts to refer to criteria such as the existence, non-existence, or relevance of the enactment, by the persons (potentially) responsible for the damage caused to consumers, of corrective actions, which needed to be taken in order to enhance the AI systems compliance with the conformity prerequisites, based on the provisions of Art. 16, let. g) and of Art. 21 of the Draft Regulation (EU) on AI (Artificial Intelligence Act);

(h) consecrating the possibility for courts, as mentioned in Art. 3 of the AILD Proposal, to impose on the producer, importer, or supplier of AI the disclosure of relevant evidence for settling the claims of the consumer/ claimant, even if the responsible person who would become the debtor of the mentioned procedural obligation would pose as a defendant since the evidence thus provided would facilitate the admission of the plaintiff's claims;

(i) enshrining the possibility for courts to operate with a reversible presumption of deficient AI systems' collapsing, suspected to be at the origin of the damage caused to consumers, while maintaining the possibility for the defendants to bring evidence of the objective performance bars of the AI system whose behaviour is subject to litigation.

3.2 Recourse to the presumption of causality in claims against non-professional AI users

3.2.1 Implications of the risks-benefits test for AI defectiveness

The contemporary physiognomy of specific liability for stand-alone or embedded AI defective products is considerably influenced by the 'confrontation for pre-eminence' waged between the two possible (classical) product defect assessment criteria, namely

[37] Luciano Floridi, 'AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models' (2023) 36 Philosophy and Technology <https://doi.org/10.1007/s13347-023-00621-y> accessed 12 July 2023.

the consumer's 'justified expectations' test versus the ratio of risks and benefits; thus, far from forming a true hierarchy or pre-eminence, each of these criteria provides a different role for national courts: reporting to an abstract standard while addressing the experiences of the average consumer or, conversely, taking into account typical product safety expectations and, on the other part, the engagement in the administration of the evidence from which the apparently (unreasonably) exorbitant weighting of the risks emerged in relation to the benefits of the consumption/use of the AI product. At one pole, the 'reasonable consumer expectations' test, according to which manufacturers/suppliers are liable if they have generated an embedded or autonomous AI product that is placed on the market in a defective and dangerous condition, exploits to a greater extent the criteria than would have been considered, through the lens of an ordinary degree of risk, by the average consumer who purchased the respective product; the mentioned criterion implies an assessment through the prism of the usual knowledge or the possession of common knowledge by the 'average consumer', with regard to the characteristics of the respective AI product, which implies engaging into an objective responsibility (as a form of strict liability[38], detached from the subjective element of fault / gross negligence) of manufacturers for AI products that present safety risks beyond the limit of reasonableness or beyond the safety threshold expected by consumers.

At the opposite pole lies the 'risks-benefits' test, under which the defendant's conduct is assessed by reference to the latter's capacity to have moderated or avoided harmful results by adopting reasonable alternatives (in the case of design defects when the product was designed based on technical parameters, according to which the benefits of consumption are exorbitantly exceeded by the security risks brought by the consumption of that product); the latter criterion uses the conceptual framework or the dogmatic paradigm of negligence imputable to the manufacturer, who accepted the initiation of the production process based on paradigms designed in an excessively risky manner referring to the benefits or the target utility (the so-called 'deficient AI design'). In practice, for products with digital content, for example, the courts might tend to oscillate between the two criteria, and to either intermittently adopt one of these criteria, or use both tests consecutively.

On another side of the discussion, important questions might be raised related to the findings by national courts of potential (yet non-manifested) defects of embedded or autonomous AI products introduced on the market, as a basis for engaging the specific liability of the producers/ the importers; these aspects fuel a constant aporia, targeting the emblematic issues of the specific liability of AI producers, designers or importers for potential AI defects whose manifestation varies on different time bars.

[38] Mark Geistfeld, 'Strict Products Liability 2.0: The Triumph of Judicial Reasoning over Mainstream Tort Theory' (2021) 14 (2) Journal of Tort Law 403, 467.

As mentioned in the previous sections, the notion of 'high-risk AI' has been circulated in the recent drafts of normative acts at the Union level to describe a risk that, following an *ex-ante* assessment of compliance[39], can be assessed as having the potential to intervene, causing physiological or patrimonial detrimental effects, or to be blamed for negatively impacting fundamental rights[40], when causing damage that requires judicial (remedial) solutions. Also, when assessing AI safety deficiencies, unfitness for consumption, or defectiveness of these products, it may become difficult or pernicious and inappropriate to develop synchronised standards for eclectic types of algorithms, as they require adaptation to the inherent risk valences of different types of AI products. It should be noted that, as follows from the provisions of Art. 9 para 2 of the AI Act Proposal, for products that involve autonomous/embedded AI with a high degree of potential risks for the bodily safety of users, the AI producer/supplier would be expected to implement a risk management system in the sense of continuous monitoring/intervention (as a permanent iterative process) for the entire usage time bars/exploitation cycle of the respective products.

It must also be observed that the 'fidelity' to the objective nature of specific liability for defective products has been kept intact in the text of Proposal 2021/0106 (Artificial Intelligence Act); in order to engage the producers' civil liability, it would suffice that an illegal action or omission was established (for example, the launching into circulation of an embedded AI or autonomous AI presenting design deficiencies), connected, or, at least, placed in a sufficiently characterised connection with the damage caused to the consumer, without being necessary that this action/omission was imputed to the AI producer. Thus, the plaintiff would not be expected to bring proof that the illegal action/omission was intentionally directed towards causing damageable results, since it is not necessary to prove, within the scope of the specific non-contractual liability, the gross fault, or the inexcusable negligence of the AI producer / AI importer.

3.2.2 AI design flaws, AI manufacturing flaws, AI implementing flaws and relevant taxonomies

It should be emphasised that not only strictly manufacturing defects are included in the substantial scope of application of the specific type of civil liability, which also includes design defects (so-called AI design defects and AI design inadequacy), respectively delivery deficiencies/omissions, incorrect and incomplete information offered to

[39] Gianclaudio Malgieri and Frank Pasquale, 'Licensing High-Risk AI: Towards Ex Ante Justification of a Disruptive Technology' [2023] SSNR Electronic Journal <https://ssrn.com/abstract=4346120> accessed 12 July 2023.
[40] Sara Gerke and Delaram Rezaeikhonakdar, 'Privacy Aspects of Direct-to-Consumer Artificial Intelligence / Machine Learning Health Apps' (2022) 6 Intelligence-Based Medicine <https://doi.org/10.1016/j.ibmed.2022.100061>accessed 12 July 2023.

consumers[41] (on technical characteristics and parameters, limited performance of the autonomous or embedded AI product, etc.), since all these types of deficiencies may per se represent a source of producers' liability[42]. The omission of adequate information regarding the foreseeable risks[43] is considered per se a defect of the product for which the civil liability of the manufacturer/importer may be engaged under EU regulations, particularly for physical/property damage caused to consumers through interaction with the AI defective product. Thus, the typical conceptual pillars include the malfunctioning of the AI systems that triggered the loss or compromise of the integrity of the consumer's data, which allows the consumer to recover, by way of a court action, the costs of data recovery/restoration: (i) the *stricto sensu* defectiveness, or unintended syncopation in the manufacturing process; in using these assumptions, the courts analyse the liability prerequisites by comparing the characteristics of the resulting AI product with the admissible safety standards, in an attempt to identify possible inadvertences of the resulting product; (ii) the design defects, when from the 'risks-benefits' ratio applied in these cases, it follows that, in hypotheses where the risks considerably exceed the benefits brought by the use of the AI product, diametrically opposed to the situations described as representing 'manufacturing defects' in a *stricto sensu* perception, the product obtained in the manufacturing process meets the technical parameters that the producer intended to achieve, while the manufacturer accepted the initiation of the production process based on design flaws which were initially assessed as adequately balancing the risks and benefits involved; in the interaction with the final consumer[44], the embedded or autonomous AI might appear excessively perilous for the users' physical/psychological safety; (iii) informational deficiencies regarding the instructions addressing the issues of integrating the AI product into the digital environment controlled by the consumer (where applicable); the informative omissions/erroneous or apparently incomplete information might represent grounds for retaining the producer's / importer's extra-contractual civil liability, in a similar manner to that which is used in the case of the above-mentioned categories of the AI defectiveness.

When assessing the damage caused by AI systems in interaction with consumers, it is worth emphasising the accent placed, in the AILD Proposal, on facilitating the administration of evidence regarding the existence of a connection, on the one hand, between the defectiveness of the AI system and the damage caused to the consumer and, on the other hand, of a causal link between the behaviour attributable to the

[41] Monika Namysłowska, Agnieszka Jabłonowska, Katarzyna de Lazari-Radek, 'AI-driven personalisation – a new challenge for Consumer law' in Luigi Lai and Marek Świerczyński (eds), *Legal and technical aspects of artificial intelligence* (Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego 2021) 95, 114.
[42] Olivier Musy and Bertrand Chopard, 'Market for Artificial Intelligence in Health Care and Compensation for Medical Errors' (2023) 75 (C) International Review of Law and Economics.
[43] Monica Navarro-Michel, 'Vehículos automatizados y responsabilidad por producto defectuoso' (2020) 7 (5) Revista de Derecho Civil 175, 223.
[44] Carlos Gómez Ligüerre and Tomás Gabriel García-Micó, 'Responsabilidad por daños causados por la Inteligencia Artificial y otras tecnologías emergentes' (2020) 1 InDret 501.

producer/importer or the supplier of the deficient AI system and the prejudice triggered for the claimant, as an aspect the importance of which becomes more visible in the light of the AI producer's/importer's failure to comply with an obligation of diligence. Encapsulated in the efforts of recognising the inherent procedural challenges in assessing the inadequacy of AI to acceptable safety standards[45] and, on the other side, the harmful results /prejudicial effects of the AI systems that generated relevant damages/harmful effects, the reversible presumption of causality integrated in the text of Art. 4 para 1 of the AILD Proposal would represent per se a significant elevator of consumers' chances in accessing pertinent evidence in damageable AI-related litigious contexts[46].

As mentioned in the previous paragraphs, in order to preserve the prerequisites of establishing the fault of the responsible persons, the defective behaviour of the AI systems might be linked to a possible failure to comply with the 'duty of care' placed on the AI producer/importer or on the retail distributor. It remains essential to note that the said failure to comply with a duty of care is expected to be assessed either in accordance with the provisions of the future Regulation (EU) on AI (i) or in accordance with other rules established at the EU level, namely those sets of nomothetic provisions that regulate the obligation of automatic and continuous monitoring of the behaviour of AI products launched on the market, in interactions with consumers or those that regulate, for example, the operation of unmanned aircraft or public means of transport without a human operator (ii).

3.3 Dichotomic approaches to the administration of evidence in liability cases concerning medium/low risk AI systems

In the background of the discussion, when engaging in the assessing of the elementary premises for maintaining the specific liability of the AI system producer/importer, although the presumption of causation gains in relevance when it can be considered sufficiently probable that the omission or action of the manufacturer/designer decisively influenced the deficient performance of the AI system, the claimant is expected to produce evidence of the existence of the prejudice. Faced, on the other hand, with the need to provide consistent evidence that the result generated by the medium /low-risk AI system has damageable effects, the consumer claimant remains required to bring evidence on the causal relationship between the claimed damage and the alleged malfunction of the system AI, at least in the perimeter of assessing the 'behaviour' of AI systems in the medium /low-risk categories; only in the case of "high risk" AI systems, as stated in the text of Art. 4 para 4 of the AILD Proposal, the courts might resort to

[45] Philipp Hacker, 'The European AI liability directives – Critique of a half-hearted approach and lessons for the future' (2023) 51 Computer Law and Security Review <https://doi.org/10.1016/j.clsr.2023.105871> accessed 16 Sept. 2023.
[46] Philipp Hacker, Ralf Krestel, Stefan Grundmann, Felix Naumann, 'Explainable AI under contract and tort law: legal incentives and technical challenges' (2020) 28 Artificial Intelligence and Law 415, 439 <https://doi.org/10.1007/s10506-020-09260-6> accessed 12 July 2023.

establishing an exception from the need for the consumer to establish the elements of causality; nevertheless, the reversible presumption of causation remains significantly functional, except in cases where the professional defendant demonstrates that, due to the absent recourse to the presumption of causation, sufficient evidence and an appropriate level of expertise were reasonably available to the consumer to prove 'on its own merits' the interrelatedness between the AI system's malfunction and the bodily injury/ property damage registered by the plaintiff.

The controversies are fuelled by the fact that the incidence of the reversible presumption is subject to the decision of the court invested in the settlement of the dispute, which may find that it would be excessively difficult for the plaintiff to provide evidence regarding the causation of the damage should no presumed interconnection be established. Evidently, other evidentiary difficulties will be assessed considering the characteristics of AI technologies, in terms of autonomy and opacity, which make explaining the inner workings of the AI system almost impossible from the perspective of the evidence available to the average consumer, the latter being (almost invariably) incapable of proving (without resorting to a reversible presumption) a causal link between the defendant's decision to launch the defective AI product on the market and the damage caused to users in the interaction with the AI system (as specified in Art. 4 para 7 of the AILD Proposal); the AI manufacturer/importer will be able to try to overturn the relative presumption of causation, while also having the possibility to invoke *ab initio* the elimination of the incidental resumption of causation, on the grounds that the circumstances of the case do not require recourse to 'invasive' procedural means.

In terms of the proportionality requirements for ordering the disclosure or preservation (by the defendant AI manufacturer/supplier) of evidence regarding the compliance/ dysfunctionality of "high-risk" AI systems, it is worth noting that the assessment of the plausibility of the claim remains essential within the scope of the analysis undertaken a priori by the courts, when ordering the defendant to disclose relevant evidence regarding the behaviour of suspected "high-risk" AI systems, even if this evidence incriminates the defendant, distancing from the procedural mechanism of *actor incumbit probatio* (given the fact that the lay consumer does not, most often, possess conclusive or relevant evidence regarding the placement of the dysfunctionalities of the AI system at the time of its launching into circulation). The said mission of the national court may come into tension with the requirements to consider the legitimate interests/fundamental rights[47] and confidential information (particularly information related to general security when interacting with AI systems). Recurring cyclically in the perimeter of the assessment of proportionality, the issue of respecting the legitimate interests of third parties is (jointly) related to the provision of adequate procedural guarantees against which the defendant,

[47] Bartłomiej Oręziak, 'Artificial intelligence and human rights' in Luigi Lai and Marek Świerczyński (eds), *Legal and technical aspects of artificial intelligence* (Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego 2021) 217, 230.

that was ordered to disclose or preserve the evidence regarding AI deficiencies, may challenge, under procedural terms, the legitimacy of evidence disclosure measures ordered by the court; the central question of the plausibility of the request encapsulates the need, for the national courts, to establish the existence of sufficient reasons for ordering evidence disclosure regarding the conduct of "high-risk" AI systems suspected of causing harmful effects.

In an aporia under which there is retaliatory bias regarding the delimitation of the sphere of responsible persons, it is important to note that the manufacturer/importer of products in the autonomous/incorporated AI categories will, in principle, be considered responsible for prejudicial reverberations, the solution mainly exploiting conceptual pillars such as: (i) the manufacturer is expected to exercise control over the selecting of technical procedures, compatibility with cyber security requirements in the design of the AI systems; (ii) when complying with the 'duty of care', the AI manufacturer/importer is expected to assume the risks resulting from the performance of these activities, to the extent that the manufacturer/designer and/or importer of the product (in the autonomous/incorporated defective AI categories) cannot resort to liability-exonerating contractual terms. On the other side, as it follows from the 'duty of care' obligation and responsibility related to the use of artificial intelligence technologies, liability effects are expected to fall on the actors most able to control certain specific risks; additionally, important attention must be paid to the fact that AI system providers within the EU's territorial settings are subject to the EU' regulatory framework, regardless of where the factual AI operator was based.

In particular, excessively risky AI design would represent legal grounds (in 'risks-benefits test failure' scenarios) for the specific liability of the manufacturer/designer/importer/supplier of a "high-risk AI system" for physical harm/ property harm caused to consumers, in each case where: (a) the harm was related to the use of an AI system that involved training models that would not meet the inherent safety criteria; (b) the courts would have consistent reasons to retain justified criticism regarding the manner under which the AI system was inconsistent with the transparency requirements provided for in the text of Art. 13 of the drafted (EU) AI Act; (c) the court would recognise the existence of some deficient characteristics of the AI system designed and developed by the responsible persons through the lens of the typical use for which the AI product was intended or through the lens of the typical purpose found in the absence of an adequate level of accuracy, robustness and cyber security as mentioned in Art. 15 and 16 of the Draft Regulation (EU) on artificial intelligence; (d) the manufacturer/importer did not immediately take relevant corrective actions, as applicable.

Obliquely mirrored in the structural requirements of liability, as pointed out, due to the complexity of the technologies involved, the element that emerges from the wording of the mentioned texts refers to the fact that it remains a challenge for the injured

consumer to correctly and pertinently identify the responsible 'AI producer', particularly in cases where the provider that 'trained' an AI application may be distinct from the original manufacturer and may have made a safety error that caused the harmful outcome in terms of bodily injury, biological harm, or patrimonial losses. B2C service providers using algorithmic systems[48] or other AI devices may face civil liability for negligent behaviour, especially when the latter neglect safety measures or instructions recommended by the manufacturer of the AI product[49].

The subrogation in the compensatory rights and the adjustment of the personal sphere of incidence of the extra-contractual civil liability for AI conduct would also represent significant progress, since recognising the quality of the "plaintiff" would not only be possible for the individual consumer who submits a claim for compensation, when a bodily/property damage was generated as a result of the action/omission of an AI system, but also for the person who was subrogated in the victim's compensatory rights or acting based on a mandate of joint representation, against the manufacturer / importer of the AI system or its authorised representative, as resulting from Art. 2 pt. (6) of the Proposal for Regulation (EU) on artificial intelligence. Similarly, in cases where the manufacturing defect of the AI product coexisted with elements related to the culpable act, unjustifiable omission, or culpable conduct[50] of a third party, and the manufacturer was held responsible for financial coverage of the damage caused to the consumer, the AI manufacturer/importer subsequently may pose as a plaintiff in a regress action against other actors involved upstream in the AI systems design/production/distribution chain.

Deciphered as 'central' or at least decisively prevalent in the liability regime of manufacturers/designers/importers of defective AI, the ordering (by the court) of the obligation (for the defendant) to make available to the court the evidence from which it would result in the involvement of the AI system in generating the damage invoked by the complaining consumer remains dependent on the requirements of proportionality; the granting of an order against the manufacturer to allow access to the relevant incriminatory evidence remains dependent on the outcome of the 'proportionality test', particularly in contexts where the 'disclosure order' was obtained against relevant third-party AI providers who are procedurally intervening in the dispute between the aggrieved consumer and the AI manufacturer / AI importer. When discussing the presence of recurring or, on the contrary, innovative aspects of AI-related specific liability, it becomes crucial to highlight how the involvement of autonomous/embedded AI systems can affect

[48] Ugo Pagallo, Marcelo Corrales, Mark Fenwick, Nikolaus Forgó, 'The Rise of Robotics & AI: Technological Advances & Normative Dilemmas' in Marcelo Corrales, Mark Fenwick, Nikolaus Forgó (eds), *Robotics, AI and the Future of Law* (Springer 2018).

[49] Paulo Henrique Padovan, Clarice Marinho Martins, Chris Reed, 'Black is the new orange: how to determine AI liability' (2023) 31 Artificial Intelligence and Law 133, 167 <https://doi.org/10.1007/s10506-022-09308-9> accessed 12 July 2023.

[50] Carsten Orwat, Jascha Bareis, Anja Folberth, Jutta Jahnel, Christian Wadephul, 'Normative Challenges of Risk Regulation of Artificial Intelligence and Automated Decision-Making' [2022] KIT Scientific Working Papers 197 <http://dx.doi.org/10.2139/ssrn.4274828> accessed 12 July 2023.

the assessment of the classical terms[51] of liability for damages caused to users, such as be the requirements relative to the causal relationship and the fault of the defendant professional; the premises of establishing liability for damages caused by defective AI products were strongly influenced by its dissociation from the element of fault or the concept of 'culpable actions/omissions' of the responsible persons; however, proof of a manufacturing defect related to the behaviour of the AI product in interaction with users remains necessary, conditional on establishing manufacturer/importer liability. In a conjugate way, the pair of relative presumptions, namely the presumption of causation between the damage caused to the consumer and the culpable behaviour of the AI manufacturer / designer, respectively the presumption of causation between the recorded damage and the defect manifested in the operating of the AI system represents a major step towards facilitating the administration of evidence or incriminating elements.

It is also remarkable that, in the AILD Proposal, no detailed attention was paid to the problem arising from establishing whether, for the assessment of AI product failure, the 'risks-benefits' ratio could present procedural pertinency or, on the contrary, adequate results could be obtained by using the test of consumers' legitimate (reasonable) expectations. The dilemmatic atmosphere of this regime of legal liability is completed by the fact that the pivotal element remains the condition of the existence of a manufacturing/design defect of the AI systems; obviously, the regime of extra-contractual liability will require a nuanced approach to the typology of covered defects, which includes their design deficiencies, regardless of whether a manufacturing defect, understood as syncope in the production process, was involved in causing the damage to the final consumer or, on the contrary, a design error or overestimation of the cyber security offered by the implementation of the respective AI system was involved, including informational deficiencies seen through the lens of the impossibility of recovering the compromised data[52], as a type of autonomous damage caused to consumers.

Another salient aspect is that, as resulting from Art. 4 para 2 of the AILD Proposal, the courts, in the event of a dispute having as its object the engagement of extra-contractual civil liability of the manufacturer, supplier or importer of the AI systems, would be able to assess the pertinency of the measures taken by the responsible persons within the applicability of internal protocols, and the results of these measures in the sense of 'tempering' the harmful effects for consumers[53]. Relevant for enumerating, in the text of Art. 4 para 3 of the AILD Proposal, the aggravating circumstances that could form the argumentation in front of the court invested with the settlement of the dispute that

[51] Jaap Hage, 'Theoretical foundations for the responsibility of autonomous agents' (2017) 25 Artificial Intelligence and Law 255, 271 <https://doi.org/10.1007/s10506-017-9208-7> accessed 12 July 2023.
[52] Leon Yehuda Anidjar, Nizan Geslevich Packin, Argyri Panezi, 'The Matrix of Privacy: Data Infrastructure in the AI-Powered Metaverse' (Faculty of Law Blogs, University of Oxford, 5 April 2023) <https://blogs.law.ox.ac.uk/oblb/blog-post/2023/04/matrix-privacy-data-infrastructure-ai-powered-metaverse> accessed 12 July 2023.
[53] Stefan Heiss, 'Artificial Intelligence Meets European Union Law' (2021) 10 (6) Journal of European Consumer and Market Law 252, 256.

justifies the retaining of extra-contractual civil liability for the defectiveness of the AI system, remains the claimant's capacity to reasonably access sufficient evidence and expertise to establish the existence of AI harmful conduct. From the taxonomy mentioned in the text of the normative act, it follows that this relative presumption of causality, although it may turn into a procedural reflex of the courts, would only be used in cases in which, when facing the opacity and accentuated complexity of the AI systems whose defectiveness was addressed, the consumer claimant would be deprived (due to the circumstances) of the real and effective possibility of procuring conclusive evidence for establishing the causal relationship connecting the defendant's culpable behaviour and the undesirable result produced by the AI system, thus generating a procedural vulnerability that the court can mitigate by resorting to the reversible presumption of causation.

It remains important to specify that the reversible presumption of causality referred to in the text of Art. 4 para 1 of the AILD Proposal would be plainly inapplicable in civil liability actions referring to AI systems in the medium risk/low-risk categories, and that the discussed remedial measures would become incidental only to the extent that the courts would consider it to present procedural pertinency, since, as stated *expressis verbis* in the text of Art. 4 para 5 of the AILD Proposal, in cases of requests for compensation related to the prejudicial conducts of medium-risk AI systems, the rebuttable presumption of causality would become applicable only if the national courts assess that it would be considerably pernicious for the plaintiff to attempt to prove the causality link through personal efforts. Concluding, it can be stated that, within the scope of civil liability actions for damages caused to the consumer by interaction with AI systems in the "high risks" category, the reversible presumption of causality referred to in the text of Art. 4 para 1 of the AILD Proposal represents the new procedural main option, the courts being expected to waive its application only to the extent relevant evidence would suffice in the absence of the 'rebuttable presumption' mechanism; on the contrary, in the perimeter of civil liability actions requesting for remedial measures for damages caused to consumers by the interaction with AI systems from the 'medium risks' or 'low risks' categories, the incidence of the reversible presumption of causation remains exceptional, intervening only to the extent that the courts would resort to the discussed presumption in order to strengthen the procedural position of the plaintiff who would not be presented the possibility of accessing conclusive evidence in the sense of establishing the originating of the alleged damage in the actions/omissions of the AI systems producer/importer or supplier.

The assumptions set out in Art. 4 para 6 of the AILD Proposal are related to the situations in which the defendant (the potential person responsible for covering the damage caused to the final consumer) is, in turn, a supplier who exploits the AI system for non-professional purposes, in which case the presumption provided for in para. (1) applies only if the defendant significantly intervened in the operating conditions, yet it

lacked pertinent measures oriented towards damage avoidance. The reversible, non-irrefragable nature of the discussed presumption of causality is expressly enshrined in the text of Art. 4 para 7 of the AILD Proposal, since in actions directed against AI suppliers/producers for the damages caused to consumers in interaction with AI systems regimented in the "high risks" categories (and, even more so, for those in the medium risks/low risks categories), the courts would be expected to allow the defendant to overturn the rebuttable presumption, the latter having the possibility to propose to the court the administration of evidence in this regard.

Innovative, in terms of the prerogatives that the court can dispose of in civil liability actions directed against AI suppliers/producers for the damages caused to the consumer by the interaction with deficient AI systems, the provisions of Art. 3 paras 1-4 of the AILD Proposal would enable national courts, either based on Art. 24 and Art. 28 para 1 of Regulation (EU) on AI, or on the AI user, to compel the defendant (the manufacturer, designer, or supplier of the AI system) to disclose the relevant evidence at the court's disposal regarding a certain system of "high-risk" AI that is suspected of causing harm to the complaining consumer. As resulting from Art. 3 para 1 of the AILD Proposal, it follows that the mentioned possibility, as a prerogative of the courts in managing the administration of conclusive evidence, remains confined to the perimeter described by the principle of proportionality of judicial measures taken during the trial, by referring, at the same time, to the plausibility (in a summary analysis, prima facie, of the relevance of the object of the requests made by the consumer/claimant); thus, proportionality requirements remain decisive for ordering (by the court) the disclosure or preservation (by the defendant AI manufacturer/ supplier) of evidence regarding the compliance/ malfunction of high-risk AI systems. The assessment of the plausibility of the claim for compensation remains essential within the scope of the analysis undertaken *a priori* by the courts[54] that may order the defendant to disclose the relevant evidence regarding the conduct of "high-risk" AI systems, even if this evidence incriminates the AI supplier[55]. The procedural context is characterised by the fact that the consumer would lack access to relevant evidence regarding the malfunctions of the AI system potentially identifiable at the time of its launching into circulation.

Additionally, the principle of moderating consumers' procedural disadvantage would be applied, as a distinct obligation incumbent on the defendant in such situations, since from the provisions of Art. 3 para 2 of the AILD Proposal, it follows that, when assessing the pertinency of a request for compensation, the national courts would resort to ordering evidence disclosure should the plaintiff have made all reasonable attempts to collect the

[54] Stanley Greenstein, 'Preserving the rule of law in the era of artificial intelligence (AI)' (2022) 30 Artificial Intelligence and Law 291, 323 <https://doi.org/10.1007/s10506-021-09294-4> accessed 12 July 2023.
[55] Regine Paul, 'The Politics of Regulating Artificial Intelligence Technologies: A Competition State Perspective' in Regine Paul, Emma Carmel, Jennifer Cobbe (eds), *Handbook on Public Policy and Artificial Intelligence* (Cheltenham: Edward Elgar forthcoming) <https://ssrn.com/abstract=4272867> accessed 12 July 2023.

relevant evidence from the defendants; nevertheless, when the quantitative or qualitative inadequacy of the evidence proposed by the plaintiff is due to the latter's inexcusable passivity, rather than the opacity or complexity of the AI system whose defectiveness is invoked, the defendant may challenge the pertinency of an evidence disclosure order.

Concerning the preventive procedural measures, as referred to in Art. 3 para 3 of the AILD Proposal, it is worth noting that, in order to prevent situations in which the defendant could be tempted by the idea of destroying the relevant evidence connected to the dysfunctionality of the litigious AI system, evidence that indicts the defendant, the court may order the defendant to preserve conclusive data and information for the period of time set by the court, with the prohibition to delete or destroy the data/information the defendant controls; this procedural measure might be ordered autonomously or simultaneously with obliging the defendant to disclose to the court / to 'declassify' relevant information regarding the security deficiencies / non-compliance of the AI systems invoked by the plaintiff.

As in the case of the first reversible presumption discussed (relating to the interconnectedness of the defendant's culpable conduct and the undesirable result produced by the AI system), the second relative presumption regarding the defendant professional's failure to comply with the requirements of the 'duty of care' (irrespectively of ordering the mandatory disclosure of relevant inculpatory evidence), the principle of proportionality of the ordered judicial measure remains applicable. As noted in Art. 3 para 4 of the AILD Proposal, national courts are expected to limit the issuing of a 'disclosure of evidence' order to cases where the 'proportionality test' has been passed by the proposed procedural measures.

As follows from the text of the mentioned normative act, by ordering the defendant IA manufacturer/ supplier to 'declassify' relevant information regarding the security deficiencies / non-compliance of AI systems, the courts are expected to 'calibrate' the mentioned procedural measure according to (i) the desiderata of preserving the commercial interests[56] of third parties involved directly or indirectly in complying with the measure ordered, especially in terms of the need to protect commercial secrecy or the autonomous rights of third parties; (ii) the concurrent recognition of the defendant's procedural prerogatives to challenge the fairness or appropriateness of the judicial measure ordered against the latter.

Similarly, the following elements are relevant for suitably understanding the cited provisions: (i) the reversible nature of the presumption of culpable failure to comply with the requirements of the duty of care implies, for the defendant involved, the possibility of overturning the presumption of inexcusable negligence invoked against the latter, while bringing evidence based on which the court would be able to assess the level of vigilance[57]

[56] Mehmet Fatih Burak (n 9).
[57] Maarten Herbosch, 'The Diligent Use of AI Systems: A Risk Worth Taking?' (2022) 11 (1) Journal of European Consumer and Market Law 14, 22.

manifested by the AI producer or AI supplier regarding security incidents; (ii) at the antipode of the presumption of good faith circulated in a litigious context by reference to the common regime of civil responsibility in most of the Member-States, in the perimeter of retaining AI manufacturers'/suppliers' liability in relation with the injured consumer, the courts would be able to operate with a relative presumption of faulty conduct applied to the recalcitrant behaviour of the defendant against whom the court ordered the obligation to divulge or preserve the evidence at the court's disposal pursuant to the issuing of a 'disclosure of evidence' order that the defendant deliberately chose to ignore.

It should be noted that this autonomous category of material damages consisting of the involuntary alteration/ compromise / deletion of data uploaded by / for the consumer does not include the compensations that could be separately requested from the data operators, for damages related to non-compliance to the General Regulation (EU) 2016/679 on data protection or to the ePrivacy Directive, as these rights of the data subjects remain the subject of separate actions without being affected by the possible introduction of liability actions directed towards recovering the consumer's data restoring costs (as expenses included in the category of damages recoverable from the defendant manufacturer/ supplier to whom the shortcomings of the AI systems were attributable, or the dysfunction of which was at the origin of the compromising of consumer's data).

The 'hybrid' nature of civil liability 'adapted' to prejudicial situations involving AI responses/omissions due to design/manufacturing defects is fuelled by the necessity of preserving the subjective element of the fault of the responsible persons, while resorting to a relative presumption of causality applied against the manufacturer/importer or the supplier of the defective AI, at the opposite of the presumption of good faith characteristic of the classic regimes of civil liability (i), seconded by the implications of a relative presumption of causality between the illegal action of launching the defective AI system on the market[58] and the existence of the design/ manufacturing defect and bodily/property damage caused to the consumer[59]. The considerably more 'fluid' approach to the subjective element of the supplier of defective AI's responsibility would depend on the 'classic' liability regimen to absorb procedural mechanisms involving a rebuttable presumption on the existence of imputable behaviour, that would allow the plaintiffs to engage in remedial actions without being placed in the position of facing insurmountable difficulties, particularly in complying to evidentiary requirements regarding the AI's (opaque[60]) behaviour, at the source of bodily/property damages. Thus, it would suffice for the plaintiff to prove the existence of the claimed damage, while

[58] Geraint Howells, Christian Twigg-Flesner (n 23).
[59] Christoph Busch, 'Rethinking Product Liability Rules for Online Marketplaces: A Comparative Perspective' (The 49th Research Conference on Communication, Information and Internet Policy, 22-24 September 2021) <https://ssrn.com/abstract=3897602> accessed 12 July 2023.
[60] Henry L. Fraser, Rhyle Simcock, Aaron J. Snoswell, 'AI Opacity and Explainability in Tort Litigation' (2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT 2022), Seoul Republic of Korea, June 21 - 24, 2022) <https://dl.acm.org/doi/10.1145/3531146.3533084> accessed 12 July 2023.

allowing those responsible for designing/manufacturing or distributing the deficient AI the chance to overturn the reversible presumption of causation, starting from the premise that the relevant information[61] regarding the (non)causality nexus is rather exclusively in the latter's possession, than in the possession of the profane consumer.

These conclusions are reinforced by the specification, in the text of Art. 4 para 2 let. (a) of the AILD Proposal, of the possibility, for the courts, to refer to the existence/non-existence and degree of adequacy of the procedural measures in the assessment of the conduct of the responsible persons (AI producer or supplier) within the framework of the AI supplier's internal security-assessment protocols and the results achieved; at the same time, the courts remain free to assess the existence of the illegal action/omission of the responsible persons by referring to compliance with the transparency obligation incumbent on AI providers, as stated in Art. 13 of the Draft Regulation (EU) on AI. Therefore, the requirement of fault-proving is not necessarily enumerated among the conditions for engaging the liability of manufacturers/importers of defective AI products, except when the consumer accessed specific remedies for civil liability (in principle, non-contractual remedies) founded on the subjective element of the responsible person's faulty conduct; the existence of damage the extent of which can be proven by the complaining consumer remains indispensable in the perimeter of the specific types of civil liability for AI defectiveness, with the specification that the biological damages are also retained when transposed into a pecuniary equivalent by the courts.

Considering the fact that, in the hypothesis that the damage caused to the consumer by the defectiveness of the incriminated IA system is bodily damage, its compensation is difficult, in view of its uncertain or even fluctuating contours over time, the evolving nature of the value claim having as its owner[62] the consumer will involve an assessment of the damage at the time of the issuing of the court's decision while taking into account the established pillars in the scope of civil liability[63] (including the missing opportunities whose materialisation would have been foreseeable for the victim of future damage or in the form of diminishing the possibilities of valorising the current existential potential, for the injured consumer[64]). Generically, this type of damage can be divided into three levels, considering, as a rule, that it consists of: (i) the economic components related to the bodily damage[65], pecuniary assessable elements (which include medical expenses and loss

[61] Cristina Frattone, 'Reasonable AI and Other Creatures. What Role for AI Standards in Liability Litigation?' (2022) 1 (3) Journal of Law, Market & Innovation <https://www.ojs.unito.it/index.php/JLMI/article/view/7166> accessed 12 July 2023.

[62] Akanksha Bisoyi, 'Ownership, liability, patentability, and creativity issues in artificial intelligence' (2022) 31 (4) Information Security Journal: A Global Perspective 377, 386 <https://doi.org/10.1080/19393555.2022.2060879> accessed 12 July 2023.

[63] David Bomhard and Marieke Merkle, 'Regulation of Artificial Intelligence' (2021) 10 (6) Journal of European Consumer and Market Law 257, 261.

[64] Bartosz Brożek and Marek Jakubiec, 'On the legal responsibility of autonomous machines' (2017) 25 Artificial Intelligence and Law 293, 304 <https://doi.org/10.1007/s10506-017-9207-8> accessed 12 July 2023.

[65] Mark Anthony Camilleri and Ciro Troise, 'Live Support by Chatbots with Artificial Intelligence: A Future Research Agenda' (2023) 17 (1) Service Business 61, 80.

of income from lucrative activities), (ii) non-economic bodily harm (involving recreational harm, generated by the physical and mental suffering of the victim, direct/indirect psychological discomfort, etc.), (iii) patrimonial harm[66] consisting in the loss/alteration of consumer data, as due to deficient AI response/ omission of an adequate response of the AI system involved. Correlatively, it remains to be noted that, regarding the non-patrimonial damages related to the interaction with AI defective products, in a deductive reasoning, the assessment of damages for psychological incidents would be particularly difficult for national courts, when connected to the bodily harm suffered by the complaining consumer[67].

The requirement of the existence of a design / manufacturing defect affecting the functionality or responsiveness of the AI system involved in generating the damage refers to a defect that endangers the safety of the consumer and/or other active components (other than the defective AI product), in an approach that draws the lines of demarcation by reference to the notion of 'design defects' of the AI system. It can be noted that the latter includes, in the perimeter of the specific liability referred to in the drafted (EU) Artificial Intelligence Act, both the intrinsic and the extrinsic defectiveness of the product (particularly for the self-learning autonomous/embedded AI categories[68]), resulting from a syncope in the manufacturing processes, either from the use in the manufacturing process of excessively risky design parameters, by reference to the product's benefits, or from deficiencies in the level of adequate and complete consumer information (non-compliance to the requirements of the obligation of transparency, including the *ex-ante* evaluation of the conformity of AI systems); on the other side, the abnormal dangerousness of the AI product launched into circulation might be assessed based on elements resulting from the fact that it does not meet the safety standards that correspond to the legitimate expectations of the consumer, or by referring to the typical characteristics of the product in autonomous AI / embedded AI category.

[66] John Linarelli, 'Artificial Intelligence and Contract Formation: Back to Contract as Bargain?' in Stacy-Ann Elvy and Nancy Kim (eds), *Emerging Issues at the Intersection of Commercial Law and Technology* (Cambridge University Press forthcoming) <https://ssrn.com/abstract=4363410> accessed 12 July 2023.
[67] Chiara Gallese, 'Suggestions for a revision of the European smart robot liability regime' in Paul Griffiths and Caroline Stockman (eds), *Highlights in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation. The PAAMS Collection* (ACPIL 2022) 29, 35.
[68] Blair Attard-Frost, 'Generative AI Systems: Impacts on Artists & Creators and Related Gaps in the Artificial Intelligence and Data Act' [2023] SSRN Electronic Journal <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4468637> accessed 12 July 2023.

4 Courts' ordering on the disclosure of relevant information on AI flaws and deficiencies

4.1 Proportionality of the judicial measure

The embodiment of the national courts' mission when deciding on the ordering of evidence disclosure may come into tension with the requirements to consider the legitimate interests/fundamental rights of the third parties concerned[69], especially the protection of trade secrets and confidential information. Recurring cyclically within the scope of the proportionality assessment, the issue of assessing the legitimate interests of third parties is (jointly) related to providing adequate procedural safeguards enabling the defendant, against whom an evidence disclosure was emitted (or who was ordered to preserve evidence regarding AI deficiencies), to challenge, procedurally, the legitimacy of measures ordered by the court in the perimeter of administrating relevant evidence.

Deliberate ignoring an evidence disclosure order constitutes legal premises for entailing the mechanism of the presumption of violation of the transparency obligations incumbent on the defendant; as resulting from the provisions of Art. 3 para 5 of the AILD Proposal, in cases in which the defendant culpably ignores a judicial order on "evidence disclosure", a national court is entitled to presume non-compliance by the defendant with a relevant "duty of care", especially in hypotheses similar to those described in Art. 4 paras 2 and 3 of the AILD Proposal, and that the compromised evidence was potentially crucial in deciphering the implications of the awarding of compensation for tortious conduct. The defendant might present a procedural interest in reversing the rebuttable presumption of culpable behaviour. Numerous conceptual and practical elements are relevant to the proper understanding of the functioning of the mentioned procedural mechanism:

(i) the rebuttable/reversible nature of the presumption of ignoring the requirements of the obligation of due diligence implies, for the defendant involved in the AI-design selection, the possibility of overturning the presumption of culpable conduct, while bringing evidence supporting its defences based on vigilant conduct;

(ii) situated at the antipode of using the presumption of good faith conveyed in a litigious context divided by reference to the 'classical' regime of tort liability, and within the scope of employing the (extra-contractual) liability[70] of manufacturers/suppliers of deficient AI systems in relation to the injured consumer's claim for compensation[71], the court will be able to operate with a relative presumption of fault applied to the

[69] Stamatis Karnouskos, 'Symbiosis with artificial intelligence via the prism of law, robots, and society' (2021) 30 Artificial Intelligence and Law 93, 115 <https://doi.org/10.1007/s10506-021-09289-1> accessed 12 July 2023.
[70] Fabiana Di Porto, 'Algorithmic disclosure rules', (2023) 31 Artificial Intelligence and Law 13, 51 <https://doi.org/10.1007/s10506-021-09302-7> accessed 12 July 2023.
[71] F Lagioia, A Jabłonowska, R Liepina, K Drazewski, 'AI in Search of Unfairness in Consumer Contracts: The Terms of Service Landscape' (2022) 45 Journal of Consumer Policy 481, 536 <https://doi.org/10.1007/s10603-022-09520-9> accessed 25 October 2023.

recalcitrant behaviour of the defendant against whom the court ordered the obligation to divulge or keep evidence at court's disposal pursuant to Art. 4 paras 1 and 2 of the AILD Proposal; moreover, it might be (reversibly) presumed that the defendant deliberately chose to ignore the content of the obligation established by the judicial order of evidence disclosure;

(iii) it follows, from the text of the mentioned normative act, that, by ordering the AI manufacturer / AI supplier to 'declassify' relevant information regarding security deficiencies / non-compliance of AI systems, the court is expected to 'calibrate' its ordered measures according to imperatives deducted from: (a) the desiderata of preserving the integrity of commercial interests of third parties directly or indirectly involved in compliance with the ordered measure, especially in view of the need to protect the confidential trade information that the respective third parties could invoke either through the lens of the protection of the autonomous procedural rights or in a particular procedural context; (b) concurrent recognition of the defendant's procedural prerogatives to challenge the fair or appropriate nature of the judicial measures on evidence disclosure, issued against the defendant.

Reminiscent from the procedural contradictory requirements arising from the principle of proportionality in assessing informational asymmetries, aspects such as the appropriateness of the judicial measures on evidence disclosure are also saliently important, as in the case of applying the rebuttable presumption regarding the culpable conduct and the undesirable results connected to the launching of defective AI systems; similar conclusions may be extracted in the case of the second relative presumption regarding 'failure' (in bad faith/from inexcusable fault) of the defendant professional's compliance with the requirements of the duty of care, for ordering the mandatory communication by the defendant of the relevant information, as noted in Art. 3 para 4 of the AILD Proposal. Additionally, the principle of moderating plaintiff's own procedural disadvantage will become applicable, as a distinct obligation incumbent on the consumer/claimant in these hypotheses, as resulting from the provisions of Art. 3 para 2 of the AILD Proposal; when the quantitative or qualitative inadequacy of the evidence proposed by the consumer/claimant is due to the latter's own inexcusable passivity (rather than to the opacity or complexity of the AI system whose defectiveness is invoked), the courts are expected to reject the request for issuing an evidence disclosure order against the defendant. From Art. 3 para 1 of the AILD Proposal, it follows that the mentioned possibility, as prerogative of the courts in managing the administration of conclusive evidence, remains confined to the perimeter described by the principle of proportionality of judicial measures taken during the litigious stages of evidence administration, by reference, simultaneously, to the plausibility (in a summary analysis, prima facie, on the relevance of the object of the requests made by the

consumer/claimant) or the pertinency of the plaintiff's claims[72]. The requirement of proportionality remains decisive for ordering the disclosure or preservation of evidence (by the defendant / AI manufacturer / AI provider) regarding compliance/ dysfunctionality of high-risk AI systems. Assessing the plausibility of the request for compensation remains essential within the scope of the analysis undertaken a priori by the courts of judgment that may order that the defendant was expected to disclose the relevant evidence regarding suspected "high-risk AI systems", even if this evidence incriminates the debtor of the obligation, as mentioned in the previous paragraphs while distancing from the exigencies of the *actor incumbit probatio* procedure. Thus, the procedural context is characterised by the fact that the consumer does not, most often, possess conclusive or relevant evidence placing the malfunctions of the AI system at the source of the prejudicial consequences.

Within the scope of addressing the civil liability for prejudicial effects generated by the interaction with defective AI systems included the "high-risk AI" categories, the rebuttable presumption of causality the incidence of which is addressed in the text of Art. 4 para 1 of the AILD Proposal would present pre-eminence, the courts waiving its application to the extent to which the defendant expressly requested the removal of the incidence of the presumption, on the grounds that the relevant evidence would be accessible to the complaining consumer which would render futile the use of the rebuttable presumption; on the contrary, in the perimeter of actions in civil liability for remedying damages caused to the consumer by the interaction with unsecured AI systems in the medium risk/low-risk category, the incidence of the relative presumption of causality is discussed as exceptionally available, intervening only to the extent that the court judges positively assessed its pertinency, meant to strengthen the procedural position of the plaintiff consumer who would be circumstantially granted the possibility of accessing conclusive evidence on the damageable results originating in the AI deficient response/omission to adequately respond. Exempted from the rebuttable presumption of causation between culpability and the undesirable result, upon request of the defending AI manufacturer/ importer / AI provider, remain the situations where, even in the absence of recourse to the substance of the said presumption in favour of the consumer/claimant, the latter would have access to sufficiently consistent expertise establishing the existence of the mentioned causal nexus.

[72] Ana Pošćić, Adrijana Martinović, 'Regulatory sandboxes under the Draft EU Artificial Intelligence Act: an opportunity for SMES?' (2022) 9 (2) InterEULawEast: Journal for the International and European Law, economics and market integrations 71, 117.

4.2 Temperament (concerning the disclosure of evidence order) pertaining to the consequences on third-party extra-patrimonial or commercial interests

When debating the preservation of the integrity of third parties' commercial interests the involvement of whom is directly or indirectly resulting in compliance with the ordered measure, the discussions refer to the necessity of protecting the confidential trade information that the respective third parties could resort to, both through the means of protecting the autonomous procedural rights to which the third parties might refer and of the recognition of the defendant's procedural prerogatives to challenge the appropriateness of the judicial measures on evidence disclosure. In order to consider the legitimate interests/fundamental rights[73] of the third parties concerned, especially the protection of trade secrets and confidential information[74], within the scope of the proportionality assessment, the courts would be expected to assess the legitimate interests of third parties while providing adequate procedural safeguards that would enable the defendant to challenge, procedurally, the legitimacy of measures ordered by the court in the perimeter of administrating relevant evidence. Thus, the AI supplier against whom an evidence disclosure order was emitted (or who was ordered to preserve evidence regarding AI deficiencies) might resort to challenging the pertinency of the judicial order, on the grounds of inexcusably ignoring the interests of third parties (including those of pre-existing B2B contractual arrangements).

5 Conceptual and practical interrogations on the pertinent use of presumptions of causality in cases concerning 'self-learning' / 'self-evolving' AI algorithmic categories

For the pertinent understanding of the 'concentric circles' involved in the subject of civil liability retained for AI products from the self-learning category, the evolutionary autonomy of these algorithmic systems remains central to the economy of the discussion regarding the initiation of actions against the AI provider. When 'dissipating' the nebulous possibility of invoking 'development risks', as a type of exonerating motif applicable in this matter, the text of Art. 6 para 1 let. (c) of the AILD Proposal retains *expressis verbis* that the manufacturer/importer can be held liable, in situations in which the defect manifested itself in the context of the self-evolving capacity of the AI product, specifying that an AI product might be defective based on its unpredictable self-learning capacities, which might evolve in 'maligned' AI behaviour.

[73] Dina Babushkina and Athanasios Votsis, 'Disruption, technology, and the question of (artificial) identity' (2022) 2 AI and Ethics 611, 622 <https://doi.org/10.1007/s43681-021-00110-y> accessed 12 July 2023.
[74] Alexandre Lodie, Stephanie Celis J. and Theodoros Karathanasis, 'Towards a new regime of civil liability for AI systems: comment on the European Commission's Proposals' (AI-regulation.com, 14 October 2022) <https://ai-regulation.com/eu-commission-proposals-on-ai-civil-liability/> accessed 12 July 2023.

On the opposite of the prerequisites that could be retained within the perimeter of 'classic' civil liability for the hidden defects of the goods that were subject to the sales contract, in which case the warranty owed to the buyer would cover only the pre-existing deficiencies at the time of contracting and the knowledge of which was not accessible to the buyer as a result of the examination of the goods with reasonable vigilance, different premises are to be observed in the case of the "extra-contractual liability" of the producers/suppliers of deficient AI systems, especially for those included in the self-learning or self-training category, or for AI systems that involve constant adaptation of responses as a result of 'learning' from their interaction with AI users.

At the same time, it remains crucial to note that the evaluation of the security deficiencies presented by self-training / self-learning AI systems[75] would be conjectural, by reference to the standard of a consumer's justified assumptions on AI product safety; as pertinent to assess the degree of dangerousness assumed by the interaction with an AI system whose responses (actions/omissions) had an evolutionary character (that are partially unpredictable, including for its creators), the consumer's 'legitimate expectations' test would sequentially focus on whether the manufacturer of each of the interconnected systems relinquished control over the AI product or algorithmic system[76], particularly in light of the general obligation on AI system providers to continuously monitor, as an itinerant and permanent obligation, the behaviour of the AI system in the interaction with consumers; the monitoring obligation represents a central pillar in the perimeter of civil liability for bodily/property prejudicial aspects affecting consumers by the malfunctioning of AI systems in the self-learning category.

6 Concluding remarks

The concept of the AI providers' 'culpable behaviour' might be seen as having, conjecturally, a multitude of meanings, in terms of transparency, devotion and fidelity towards the principles of precaution and preservation of consumers' safety, constancy and obedience to reasonable standards of probity. The mentioned semantic arborescence of the 'AI credibility' conceptual framework would find different forms of objectification in the national legal systems while placing rebuttable presumptions of causality in the vanguard of the efforts of adapting the 'classical' liability regimen to the provoking aspects of liability for AI systems defectiveness.

We argued that the binary premises (partially subjective, generated by the element of fault and partly objective, focused on the element of the presence of the design/manufacturing defect of the AI system) in engaging the specific liability of the

[75] Jarosław Protasiewicz and Marek Michajłowicz, 'Agile development of intelligent systems: a case study' in Luigi Lai and Marek Świerczyński (eds), *Legal and technical aspects of artificial intelligence* (Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego 2021) 297, 310.

[76] Jaromír Šavelka and Kevin D Ashley, 'Legal information retrieval for understanding statutory terms' (2022) 30 Artificial Intelligence and Law 245, 289 <https://doi.org/10.1007/s10506-021-09293-5> accessed 12 July 2023.

pernicious AI systems producers/importers might be contrasted to the versions of the subjective liability 'classical' regimen, especially from the perspective of consecrating a taxonomy of defects covered by the material scope of the liability, compartmentalised into manufacturing defects versus design defects and information deficiencies; congruently, failure to comply with the transparency obligation incumbent on the manufacturer/importer of deficient AI systems would represent *per se* legal ground for retaining the AI supplier's responsibility for covering the damageable effects.

In terms of 'expanding' the material scope of the extra-contractual civil liability of manufacturers, importers or suppliers of defective AI systems, the most important innovative elements consist in the expansion of the scope of legal protection rules in order to include patrimonial damages caused to the consumer by loss or corrupting of consumer's data, in the interaction with the defective AI system, unless the data was used exclusively for professional purposes, without excluding the hypotheses of coverage for the mixed purpose (simultaneously professional and personal) of the compromised data. According to Recital 16 para 1 of Draft Directive COM/2022/495, in principle, data restoration costs must be included in the category of damages recoverable from the defendant (AI manufacturer/supplier) to whom the shortcomings of the AI system are attributable. Apparently, it is not relevant, for the coverage of these categories of damages, where the data was stored, as it does not matter to any extent whether the defendant (the AI manufacturer/supplier) directly controlled the storage procedures for the consumer's lost data. Thus, the (EU) AI Act Proposal seems to engage in a long-run debate outlined around the controversial concept of 'ownership of data' stored by smart technologies or through algorithmic applications. The focus is not placed on the possible recognition of a genuine category of 'property rights' over the compromised data, the damage of which would be the subject of a civil liability action against the designer/producer or supplier of the deficient AI system, but on their inclusion in the category of recoverable material damage, without simultaneously and necessarily engaging in recognising consumer 'ownership' over compromised data, as a prerequisite for damage coverage. Nevertheless, when assessing the pertinency of a request for compensation, the national courts would resort to ordering evidence disclosure should the plaintiff have made all reasonable attempts to collect the relevant evidence. In cases when the quantitative or qualitative inadequacy of the evidence proposed by the plaintiff is due to the latter's inexcusable passivity, rather than the opacity or complexity of the AI system, the defectiveness of which is invoked, the defendant may challenge the pertinency of the evidence disclosure order. Finally, the rebuttable/reversible nature of the presumption of ignoring the requirements of the obligation of due diligence implies, for the defendant, the possibility of overturning the presumption of inexcusable negligence in breaching the 'duty of care' incumbent on AI providers, while bringing evidence illustrating the level of vigilance manifested during the AI design-selecting process.