Legal Requirements for Human Oversight Within the AI-Act-Proposal*

Moritz Griesel, Tizian Matschak**

Contents

I. Introduction 2

II. Legal Requirements 2

A. Legal Background 3

B. Status Quo of the Legislative Process 8

C. Key Regulatory Subjects in the Legislative Process 8

III. Multidisciplinary View on the Requirements of Human Oversight 27

A. Nuances of Human Oversight 27

B. Humans' Understanding of AI Reasoning 28

C. XAI as a Tool to Support Human Oversight 29

IV. Conclusion 30

V. Bibliography 31

* Disclaimer: The draft for this paper has been submitted on April 2nd 2024. Relevant literature and legal material available at that point in time have been included in this paper.

** Moritz Griesel is Ph.D. Candidate at the Institute for Commercial and Media Law (Prof. Dr. Andreas Wiebe, LL.M. (Virginia)) at the University of Göttingen (moritz.griesel@jura.uni-goettingen.de); Tizian Matschak is Ph.D. Candidate at the Chair of Information Security and Compliance (Prof. Dr. Simon Trang) at the University of Göttingen (tizian.matschak@uni-goettingen.de).

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

I. Introduction

In April 2021, the European Commission published the Proposal for an Artificial Intelligence Act (AI-Act-Com-P)[1], introducing a risk-based approach. The requirements for high-risk AI-systems are laid down in Title III Chapter 2 AI-Act-Com-P. Among others, Art. 14 (1) AI-Act-Com-P stipulates that these systems shall be designed and developed in such a way that natural persons can effectively oversee them during the period in which the AI-system is in use. In particular, the human supervisor must understand the AI-system's capabilities and limitations and supervise it appropriately. However, it remains unclear how these intensive regulatory requirements should be met in an application that is explicitly built to provide economic efficiency by *autonomous* decision-making.[2] This paper will take on this observation and will depict key regulatory requirements as set out by Art. 14 AI-Act-Com-P and compare them to the European Parliament's Proposal[3] (II.).

II. Legal Requirements

The legal requirements for human oversight are essentially laid down in Art. 14 AI-Act-Com-P. The AI-Act-Com-P was preceded by a series of expert opinions and statements as well as the Commission's own strategy papers[4], which will be briefly described below (A.). The legislative process was then initiated with the Commission's proposal, which was followed by amendments proposed by the Council and the European Parliament, which resulted in the Provisional Agreement (B.). In particular, the paper aims at providing an assessment of the core requirements of human oversight as layed down in the Provisional Agreement and compare them to the previous proposals on human oversight measures (C.). This will also include

[1] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts, COM(2021) 206 final.

[2] European Law Institute, 'Guiding Principles for Automated Decision-Maing in the EU' 23 <https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Innovation_Paper_on_Guiding_Principles_for_ADM_in_the_EU.pdf> accessed 10 August 2023; Gerhard Wiebe, 'Produktsicherheitsrechtliche Betrachtung Des Vorschlags Für Eine KI-Verordnung' [2022] Betriebs-Berater 899, 903.

[3] Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM(2021)0206, P9_TA(2023)0236.

[4] Wiebe (n 2).

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

ISO/IEC drafts for technical standards which are intended to fulfill the prospective requirements of the AI-Act.

A. Legal Background

The AI Act follows an human-centric approach, which was not only taken up with the AI-Act-Com-P but had already been formulated earlier in separate strategy papers such as the White Paper on Artificial Intelligence[5] and the European AI Strategy[6]. Even at this early stage, the Commission recognized a two-dimensional thrust: First, people should be enabled to take advantage of algorithmic and artificial intelligence systems, including by making their own informed decisions in the digital environment.[7] At the same time, they are to be protected from risks and harm to their health, safety, and fundamental rights. To this end, based on the expert opinion of the High-level Expert Group on Artificial Intelligence[8], the Commission formulated core requirements that were summarized under the term "Ethical Guidelines for Trustworthy AI".[9] The following chapter focuses on human oversight requirements in these official documents, which preceded the actual legislative process.

1. European Strategy for AI

As early as 2018, the Commission adopted a package of measures as part of the European AI Strategy that was intended to ensure the use of ethical and trustworthy artificial intelligence in the European Union as well as to show all stakeholders a transparent path toward a regulation.[10] Although the terms "human supervision" or "human oversight" at this point in time did not explicitly appear in the strategy, it does contain the first rudiments of these concepts. For example, AI-systems placed on the market should be required to allow humans to retrace the steps (or their basis) in

[5] White Paper on Artificial Intelligence - A European approach to excellence and trust 2020, COM(2020) 65 final.

[6] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe 2018, COM(2018) 237 final.

[7] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe (n 6) 16.

[8] High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI' <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

[9] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5) 9.

[10] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe (n 6) 2f.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

decision-making processes.[11] In doing so, the Commission recognized that with the development of corresponding systems, research in the area of the explainability of AI-systems would have to be extended.[12] Exactly this research area, which has in the meantime established itself as XAI or explainable AI, now covers (among others) the scientific basis of human oversight.[13]

2. Coordinated Plan for AI

In its Coordinated Plan for Artificial Intelligence[14], the Commission then identified areas for action in December 2018. This included increasing targeted public and private investment in AI in the following years, the preparation for socio-economic changes, and the initiation of an appropriate ethical and legal framework. While the term "human oversight" did not find its way into the text here either, the 2021 review of the coordinated plan[15] selectively clarified the Commission's ideas regarding human oversight. Accordingly, given the importance of the policy area and the need to ensure the protection of fundamental rights, law enforcement, migration, and asylum, AI applications are never to be used as stand-alone "decision makers".[16] Instead, they should be used to assist, for example by providing clues for

[11] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe (n 6) 17.

[12] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe (n 6) 17.

[13] Cf. for an overview of state-of-the-art human oversight measures Gesina Schwalbe and Bettina Finzel, 'A Comprehensive Taxonomy for Explainable Artificial Intelligence: A Systematic Survey of Surveys on Methods and Concepts' [2023] Data Mining and Knowledge Discovery <http://arxiv.org/abs/2105.07190> accessed 10 August 2023; Alejandro Barredo Arrieta and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' <http://arxiv.org/abs/1910.10045> accessed 10 August 2023; Arun Rai, 'Explainable AI: From Black Box to Glass Box' (2020) 48 Journal of the Academy of Marketing Science 137; Amina Adadi and Mohammed Berrada, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' (2018) 6 IEEE Access 52138.

[14] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Coordinated Plan on Artificial Intelligence 2018 COM/2018/795 final.

[15] Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Fostering a European approach to Artificial Intelligence COM(2021) 205 final.

[16] Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Fostering a European approach to Artificial Intelligence (n 15) 49.

investigations or assessments in a specific context, but the final decision should always rest with a human.[17] Consequently, in the medical context, AI should gain insights from the analysis of (patient) data to support diagnoses and treatments. Ultimately, however, the final decision on whether and how to treat should remain with a human physician.[18]

3. Report of the High Level Expert Group on AI

A more detailed description of the envisaged regulations on human oversight was finally provided only in the report of the High Level Expert Group on artificial intelligence set up by the European Commission.[19] Based on the differentiation between the primacy of human action and the actual considerations for human oversight, the expert group continued the human-centered approach of the European Commission. First, the self-determination of the user of an AI-system should not be able to be compromised by the AI-system.[20] Instead, AI-systems should take on a servant role, helping people "make better, more informed decisions in line with their own goals."[21] In contrast, forms of unfair manipulation, deception, oppression, or conditioning are not to be permitted. Thus, the expert group explicitly follows the expression of the right not to be subjected exclusively to automated decision-making already known from Art. 22 GDPR.[22]

The primacy of human action is also to be complemented by a differentiated system of human oversight. First of all, it is crucial that the expert group always understands human oversight as one of several components for ensuring ethical, trustworthy and human-centered AI.[23] Consequently, the requirements in individual cases must always be determined in the context of other security and control measures. It is

[17] Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Fostering a European approach to Artificial Intelligence (n 15) 40.

[18] Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Fostering a European approach to Artificial Intelligence (n 15) 40 at fn 184.

[19] High-Level Expert Group on AI (n 8).

[20] High-Level Expert Group on AI (n 8) para 64.

[21] High-Level Expert Group on AI (n 8) para 64.

[22] High-Level Expert Group on AI (n 8) para 64.

[23] High-Level Expert Group on AI (n 8) para 65.

conceivable, for example, that a deficit in human supervision can be compensated by intensive testing.[24]

In addition, the selection of concrete oversight measures and their intensity must also be determined depending on the concrete application of the AI-system as well as the associated (potential) risk.[25] In this respect, a differentiation is made between three different degrees of human oversight, which will be described in more detail below.[26]

4. White Paper on Artificial Intelligence

The guidelines drawn up by the High-Level Expert Group were finally taken up and explicitly adopted by the European Commission in the White Paper on Artificial Intelligence.[27] In terms of time, sufficient oversight measures should already be taken into account in product design and implemented over the entire life cycle of the AI-system.[28] Concrete measures and their intensity are to depend in particular on the intended use of the systems and the effects (including risks) that the use could have for the citizens and legal entities concerned. These specifications take account of the principle of proportionality, and the resulting flexibility makes it possible to impose correspondingly high requirements on human oversight in particularly risky use-case scenarios or where there is a high potential threat to fundamental rights. At the same time, it should also be considered here that the AI-systems regulated in this way are classified as high-risk AI-systems anyway. Due to the initial high-risk potential in this regard, the leeway on the legal consequences side (i.e., in selecting and determining the intensity of concrete measures) is limited.[29] Against this background, it is also not surprising that the potential measures not exhaustively listed in the White Paper all have a high intensity of intervention. Measures include, for example, that the output of the AI-system only becomes effective after it has been previously reviewed and validated by a human, or that the decision of the AI-system, although effective, is subject to ex post human review, which can revise the decision.[30] In addition, it should be possible to monitor the AI-system during operation and take corrective action in

[24] High-Level Expert Group on AI (n 8) para 65.

[25] High-Level Expert Group on AI (n 8) para 65.

[26] Cf. below at 0.

[27] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5) 9.

[28] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5) 15.

[29] Cf. below at II.C.2.ii.

[30] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5) 21.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

real-time or even disable it.[31] Finally, the Commission counts as human oversight the possibility of imposing system-immanent limits or operational constraints on the AI-system already at the design or production level, upon the occurrence of which the AI-system would, for example, hand over decision-making to a human.[32]

5. Right to Human Oversight, Declaration No. 9 Lit. C on European Digital Rights and Principles

In February 2023, the Parliament, the Council and the Commission issued a Declaration on the Digital Rights and Principles of the Digital Decade[33], in which the importance of human supervision was once again emphasized. The declaration is to be understood as a political commitment by the EU and its member states but does not have any further legal effect beyond this declaration of intent. While the significance of such declarations can therefore be doubted in principle[34], they should at least be clearly formulated and go beyond existing declarations of intent and policy papers. In the case of this declaration, this is only the case to a limited extent. Similar to the previously discussed strategy papers of the Commission, a human-centered approach is also pursued here. Once again, humans are to be empowered to make their own informed decisions and the influence of artificial intelligence is to be reduced to a desirable level. According to Declaration No. 9 lit. c, human oversight should consist of monitoring all outcomes that affect the safety and fundamental rights of people. However, this comprehensive approach is limited with regard to Declaration No. 9 lit. e by again referring to the implementation of proportionate measures. It is in particular this criterion of proportionality which has been decisive in the legislative process.[35]

The analysis of these soft law instruments shows that HO measures have been part of the development of a human-centered approach to AI in the EU from the very beginning. The White Paper on AI in particular contributed a great deal to the

[31] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5) 21.

[32] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5) 21.

[33] 'European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future' (7 February 2023) <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles> accessed 9 August 2023.

[34] Cristina Cocito and Paul De Hert, 'The Use of Declarations by the European Commission: "Careful with That Axe, Eugene". — 'The Digital Constitutionalist' (14 March 2023) <https://digi-con.org/the-use-of-declarations-by-the-european-commission-careful-with-that-axe-eugene/> accessed 18 March 2024.

[35] Cf. below at II.C.2.ii.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

development and already contained key aspects which – as will be shown – were later adopted by the HLEG and are also reflected in the Provisional Agreement.

B. Status Quo of the Legislative Process

After the Commission's Proposal, both the French and the Czech Presidencies have proposed two amendments each concerning – among others – the requirements for human supervision.[36] However, since most of these amendments have been taken up in the EP's Proposal, they will not be included in this overview. The legislative process just passed the trialogue negotiations resulting in the Provisional Agreement.[37]

C. Key Regulatory Subjects in the Legislative Process

The debate on human oversight in general focuses on two major aspects. On the one hand, the term high-risk AI-system, which prescribes the scope of application of Art. 14 AI-Act-PA, is controversial in two respects: in addition to the specific definition of an "AI-system", it is also discussed which AI-systems should be classified as a high-risk AI-system and thus be subject to the strict set of obligations set out in Art. 8 et seqq AI-Act-PA. On the other hand, for a long time the final set of requirements for human oversight have been uncertain, because both the Council and the EP have proposed a number of amendments following the initial Commission Proposal. The paper will briefly point towards major developments on the scope of application of Art. 14 AI-Act-PA, which occurred during the legislative process, and will then turn to a more detailed analysis of the key requirements of human oversight in the AI-Act.

1. Scope of Application

The requirement of human oversight is found in Title III Chapter 2 AI-Act-PA and is thus applicable to all high-risk AI-systems. In general, the AI-Act follows a risk-based approach, differentiating between prohibited AI-systems, which pose an unacceptable risk, high risk AI-Systems, which are subject to extensive governance meanisms, and other AI-Systems interacting with natural persons, which must fulfill general transparency obligations.

---

[36] Proposal of the French Presidency on Art. 8-15 AI-Act of 13.01.2022; Proposal of the French Presidency on Art. 14 AI-Act of 04.04.2022; Proposal of the Czech Presidency on Art. 14 AI-Act of 15.07.2022; Proposal of the Czech Presidency on Art. 1-29 AI-Act of 16.09.2022.

[37] Provisional Agreement resulting from interinstitutional negotiations, Proposal for a regulation laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts 2021/0106(COD), <https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf> accessed 11 March 2024.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

Regarding human oversight measures, it must be an AI-system, to which a high risk is attributed in accordance with Art. 6 AI-Act-PA. Moreover, according to the Commission Proposal, the addressees of regulation are providers (Art. 3 no. 2 AI-Act-Com-P) and users (Art. 3 no. 4 AI-Act-Com-P). The EP's Proposal and the Provisional Agreement also address the "deployer"[38] as defined in Art. 4 No. 4 AI-Act-EP-P and AI-Act-PA.

i. AI-System

To begin with, an AI-system is defined in Art. 3 no. 1 AI-Act-Com-P as "software that is developed with one or more of the techniques and approaches listed in Annex I AI-Act-Com-P and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with". In this respect, the Commission explicitly pursues a technology-open and thus very flexible approach, focusing in particular on the essential functional characteristics of the software.[39] Furthermore, AI-systems can be designed to operate with varying degrees of autonomy and can be used independently or as a component of a product, whether the system is physically integrated into the product (embedded) or serves the function of the product without being integrated into it (non-embedded).[40] This broad legal definition was rightly criticized, as it would mean that virtually any software would fall within the scope of the AI-Act-Com-P.[41]

The EP's Proposal in turn defines an AI-system in Art. 3 no. 1 AI-Act-EP-P as a machine-based system, that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments.

The Provisional Agreement now combines both approaches by referring in Art. 3 (1) AI-Act-PA to an "AI-system" as a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to

[38] Art. 29 AI-Act-EP-P.

[39] Recital 6 AI-Act-Com-P.

[40] Recital 6 AI-Act-Com-P.

[41] Wiebe (n 2); Philipp Roos and Caspar Alexander Weitz, 'Hochrisiko-KI-Systeme Im Kommissionsentwurf Für Eine KI-Verordnung' [2021] Multimedia und Recht 844, 845; Maria Heil, 'Die Neue KI-Verordnung (E) – Regulatorische Herausforderungen Für KI-Basierte Medizinprodukte-Software' [2022] Zeitschrift für das gesamte Medizinproduktrecht 1, 4 David Bomhard and Marieke Merkle, 'Europäische KI-Verordnung' [2021] Recht Digital 276, 277.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

ii. High-Risk AI-System

To continue, Art. 14 AI-Act-Com-P is only applicable to high-risk AI-systems in order to ensure that those systems do not pose a significant adverse impact on the health, safety and fundamental rights of persons in the Union.[42] The classification as a high-risk system can be made both via Art. 6 (1) AI-Act-PA and Art. 6 (2) AI-Act-PA.

Art. 6 (1) AI-Act-PA addresses those high-risk AI-systems, where the AI-system is intended to be used as a safety component of a product or is itself a product, covered by the Union harmonization legislation listed in Annex I AI-Act-PA. According to Art. 3 no. 14 AI-Act-PA, a safety component of a product or system means a component of a product or of a system which fulfills a safety function for that product or system or the failure or malfunctioning of which endangers the health and safety of persons or property. In addition, the product must undergo a third-party conformity assessment pursuant to Art. 6 (1) lit. b AI-Act-PA. Harmonization legislation includes various European Directives and Regulations based on the "New Legislative Framework"[43] regulating European conformity testing.[44] This includes, for example, the Machinery Directive[45], the Directive on safety of toys[46], and the Regulation on medical devices.[47]

Moreover, Art. 6 (2) AI-Act-PA address “stand-alone” AI applications, because they pose a high risk of harm to the health and safety or the fundamental rights of persons,

[42] Recital 46 s. 5 AI-Act-PA.

[43] Cf. for intersections between other NFL regulations and the AI-Act-Com-P Gabriele Mazzini and Salvatore Scalzo, ‘The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts’ [2022] SSRN Electronic Journal <https://www.ssrn.com/abstract=4098809> accessed 11 August 2023; for more details on the NFL in general cf. Carsten Schucht, ‘Die Neue Architektur Im Europäischen Produktsicherheitsrecht Nach New Legislative Framework Und Alignment Package’ [2014] Europäische Zeitschrift für Wirtschaftsrecht 848.

[44] Gerald Spindler, ‘Der Vorschlag der EU-Kommission für eine Verordnung zur Regulierung der Künstlichen Intelligenz (KI-VO-E)’ [2021] Computer und Recht 361, 366.

[45] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC [2006] OJ L 157/24.

[46] Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys [2009] OJ L 170/1.

[47] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC [2017] OJ L 117/1.

taking into account both the severity of the possible harm and its probability of occurrence.[48] However, only AI-systems used in a number of specifically pre-defined areas specified in Annex III AI-Act-PA are covered. Accordingly, AI-systems that are used in the biometric identification and categorization of natural persons[49], in the management and operation of critical infrastructure[50] or in the administration of justice and democratic processes[51] fall within the scope of application.

Finally, the Commission is empowered pursuant to Art. 7, 97 AI-Act-PA to adopt delegated acts to update the list in Annex III AI-Act-PA by adding high-risk AI-systems, which pose similar risks as those known from Art. 6 (2), Annex III AI-Act-PA and provides for greater flexibility in the future.[52]

While Art. 6 AI-Act-PA remained largely unchanged in the trialogue negotiations, the fact that its requirements may still be too imprecise in individual cases was also taken into account. According to Art. 6 (5) AI-Act-PA, within 18 months of the regulation coming into force, the Commission is to issue guidelines specifying the practical implementation of Art. 6 AI-Act-PA completed by a comprehensive list of practical examples of high risk and non-high risk use cases on AI-systems.[53] In this respect, however, it remains to be seen whether this list will actually represent a significant step towards greater legal certainty. It is also questionable whether this list would need to be continuously updated due to ongoing technical progress.

iii. Addressees of the Regulation

a) Provider

According to Art. 16 lit. a AI-Act-PA, providers of high-risk AI-systems shall ensure, that their high-risk AI-systems are compliant with the requirements set out in Chapter 2 of Title III, which particularly includes the requirements for human oversight.[54] The term 'provider' is defined in Art. 3 no. 2 AI-Act-PA and means a natural or legal person, public authority, agency or other body that develops an AI-system or a general purpose AI model or that has an AI-system or a general purpose AI model

[48] Recital 52 s. 1 AI-Act-PA.

[49] No. 1 Annex III AI-Act-PA.

[50] No. 2 Annex III AI-Act-PA.

[51] No. 8 Annex III AI-Act-PA.

[52] Spindler (n 39); Wiebe (n 2) 901.

[53] Art. 6 (2c) AI-Act-PA.

[54] Roos and Weitz (n 38) 847.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

developed and places them on the market or puts the system into service under its own name or trademark, whether for payment or free of charge. Providers should also be covered regardless of whether they are established in the Union or in a third country.[55] The term "provider" has survived the trialogue-negotiations almost unchanged. By naming this addressee of the regulation, it is already clear that human oversight should not only be taken into account at user level, but also at product design level. This is because only the provider can provide for appropriate measures at this level.

Moreover, the EP's Proposal to oblige providers to ensure that natural persons who are entrusted with the human oversight of high-risk AI-systems are specifically made aware of the risk of automation or confirmation bias was not adopted in Art. 16 AI-Act-PA. However, the proposal was not completely removed, but included in a modified form in Art. 4b AI-Act-PA. Providers must now ensure, at least for their staff and other persons dealing with the operation and use of AI-systems on their behalf, that they have a sufficient level of AI literacy, which in turn refers to (among others) the awareness about the opportunities and risks of AI and possible harm it can cause.[56] This restriction makes sense, as the provider is now only liable for those persons who are active in its sphere of interest and are therefore dependent on its instructions, for example.

b) Deployer

The second addressee of the human oversight requirements is the deployer. According to Art. 29 (1) AI-Act-PA deployers of high-risk AI-systems shall take appropriate technical and organizational measures to ensure they use such systems in accordance with the instructions of use accompanying the systems. The EP wanted to introduce an even stricter duty to comply with the requirements of Art. 14 AI-Act by proposing that to the extent deployers exercise control over the high-risk AI-system, they shall implement human oversight in accordance with Art. 14 AI-Act-EP-Proposal and ensure, that the natural persons assigned to ensure human oversight are competent, properly qualified and trained, and have the necessary resources in order to ensure the effective supervision of the AI-system. However, this proposal did not prevail the trialogue-negotiations. Instead, deployers must particularly ensure that the natural persons assigned to ensure human oversight of the high-risk AI-

[55] Recital 10 AI-Act-PA.

[56] Art. 3 (bh) AI-Act-PA, the sections in Art. 3 AI-Act-PA do not yet appear to be subject to a consecutive enumeration system, so reference is meant do be made to the definition of "AI Literacy".

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

systems have the necessary competence, training and authority as well as the necessary support.[57]

Pursuant to Art. 3 no. 4 AI-Act-EP-P, deployer means any natural or legal person, public authority, agency, or other body using an AI-system under its authority except where the AI-system is used in the course of a personal non-professional activity. In this regard it will be decisive who controls and operates[58] the AI-system not merely for private purposes.[59] The term was introduced by the EP after the Commission's Proposal to impose similar obligations on the "user" of the AI-system has met criticism.[60]

c) User

Finally, the initial Commission Proposal provided that human oversight measures may also be implemented by users of the high-risk AI-system.[61] The same provision with reference to the term "user" is also found in in Art. 14 (3) lit. b) AI-Act-EP-P, the Council's mandate and the Provisional Agreement. The term was originally defined in Art. 3 no. 4 AI-Act-Com-P as "any natural or legal person, public authority, agency or other body using an AI-system under its authority, except where the AI-system is used in the course of a personal non-professional activity". Thus, the term "user" in the Commission Proposal is identical with the term "deployer" in the EP's Proposal, the Council's mandate and the Provisional Agreement. The EP changed the term "user" to "deployer" in order to ensure a somewhat clearer conceptual distinction from private internet users.[62] While the EP changed the term in Art. 3 no. 4 AI-Act-EP-P, it left Art. 14 (3) lit. b AI-Act-EP-P untouched. It is unclear whether this is an editorial error or whether the user should continue to be the addressee of the regulation. However, taking the legislative process into account, the user should now actually be understood as the deployer. This result would also be consistent with Art. 29 AI-Act-PA, which regulates the obligations of the deployer and exclusively uses this term. However, in order to avoid legal uncertainty, the final version of the

[57] Art. 29 (1a) AI-Act-PA.

[58] Recital 59 AI-Act-Com-P.

[59] Bomhard and Merkle (n 41) 278.

[60] For more details cf. below at c).

[61] Art. 14 (3) lit. b AI-Act-Com-P.

[62] Daniel Becker and Daniel Feuerstack, 'Der Neue Entwurf Des EU-Parlaments Für Eine KI-Verordnung' [2024] Multimedia und Recht 22.

text should only use those key terms that are also defined in Art. 3 of the AI Act. This is why there is a need for improvement in this respect.

d) Sufficient Regulation?

With the provider and the deployer, the AI Act places the responsibility precisely on those two entities who can most effectively ensure human oversight at either a technical or organizational level.

However, it does not specify who on the provider/deployer side should specifically perform the task of human oversight. Instead, deployers "only" have to ensure that the natural persons who carry out human oversight have the necessary competence, training and authority as well as the necessary support.[63] This does place demands on personal expertise and intervention rights. However, it remains unclear, for example, how many qualified personnel must be deployed or whether third parties can also be assigned human oversight by the deployer.

Moreover, providers shall give *due consideration* to the technical knowledge, experience, education and training to be expected by the deployer and the presumable context in which the system is intended to be used.[64] This obligation to just give *due consideration* is fairly easy to be met – apparently, it is sufficient to provide for a reasonable assessment of the nature and extent of human oversight without any obligation to investigate.[65] However, the deployer is always dependent on the accurate preliminary work of the provider: According to Art. 14 (3) AI-Act-PA, the provider must either only identify human oversight measures or also build them into the AI-system. If human oversight measures are not or incorrectly selected or not or incorrectly build in by the provider, the deployer's oversight capabilities are therefore directly restricted. Measures should therefore inevitably be adapted to the expected level of competence of the deployer - and not only in a reasonable assessment.

This limits the regulation to a technical and organizational level, which gives providers and deployers the greatest possible leeway. However, it is important to bear in mind, that Art. 14 AI-Act-PA wants to regulate high-risk AI-systems. Whether such a

[63] Art. 29 (1a) AI-Act-PA.

[64] Art. 9 (4) AI-Act-PA.

[65] Lena Enqvist, '"Human Oversight" in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15 Law, Innovation and Technology 508; Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' [2021] Computer Law Review International 97.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

flexible regulation is appropriate in a high-risk environment is more than questionable.

2. Overview: Key Regulatory Requirements for Human Oversight

The Proposals do not explicitly identify individual measures or specific tools that human oversight must include, but set requirements that those measures and tools must meet. In general, it promotes a shared responsibility between deployers and users of an AI-system: While the deployer must include tools and methods enabling the user to oversee the AI-system, the user must subsequently exercise this oversight.[66]

As a starting point, any measure must be suitable to contribute to preventing or minimizing risks to health, safety, or fundamental rights[67] that may arise when a high-risk AI-system is used as intended or under reasonably foreseeable misuse.[68] The EP's Proposal to include the environment in this enumeration has not been adopted in the trialogue negotiations.[69] The timely limitation to intended use or reasonably foreseeable misuse is conclusive: Although Art. 14 (1) AI-Act-PA seems to limit the scope of application to the period in which the AI-system is in use, the predominantly technical and organizational nature of the measures suggests that they should be taken into account already at the design level and thus even before the AI-system is used. With the requirement of a stop/panic button, the temporal scope of application comes to a logical end, because as soon as the AI-system is stopped, the supervision object is no longer active. Instead, an error analysis will be indicated in these cases, which can be carried out on the basis of the records to be maintained in accordance with Art. 13 AI-Act-PA, for example. Accordingly, the temporal scope of the record keeping provision is extended to the entire lifecycle of the AI-system.[70]

Human oversight requirements can be (roughly) divided into two categories: First, requirements are placed on the system itself, including high-risk AI-systems being designed and developed in such a way that they can be effectively supervised by natural persons.[71] Which measure, which technology or which (human-machine)

---

[66] Johann Laux, 'Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act' (2024) AI & Society 2853 (2858).

[67] Art. 14 (2) AI-Act-Com-P; Art. 14 (2), Recital 43 AI-Act PA.

[68] Art. 14 (2) AI-Act-PA.

[69] Cf. Art. 14 (2) AI-Act-EP-P.

[70] Art. 13 (1) AI-Act-PA.

[71] Art. 14 (1) AI-Act-Com-P; Art. 14 (1) 1 AI-Act-EP-P.

interface should be applied in particular is to be determined in the individual case by taking the proportionality of the possible measures into account.[72]

Second, the Proposals also address the human component, for example by including requirements for AI-literacy.[73] The following overview is based on this two-dimensional orientation of the Proposals.

i. Design Requirements and Agency

To begin with, the measures for human oversight can be either built into the high-risk AI-system by the provider before it is placed on the market or put into service[74] or – where appropriate – be identified by the provider and subsequently be implemented by the deployer.[75] This approach indicates an intended division of tasks between providers and deployers. While providers are expected to provide and implement HO measures at the design and concept level in particular, deployers become involved later on "during the operation" of the AI-system.[76] Moreover, the providers' obligation extends to any further development of their AI-systems – otherwise there would be a risk that user oversight would come to nothing.[77]

Art. 14 (4) s. 1 AI-Act-PA clarifies that human oversight can only be exercised by natural persons. This clarification was necessary because the Commission Proposal in Art. 14 (3) Alt. 2 AI-Act-Com-P still wanted to assign human oversight to the user, who could also be a legal person according to Art. 3 No. 4 AI-Act-Com-P. Moreover, Art. 14 (4) AI-Act-Com-P only contained the provision that oversight was incumbent on individuals (German text version: "Personen"), which seemed ambiguous, at least due to the differentiation between legal and natural persons. Therefore, the

[72] Cf. below at II.C.2.ii.

[73] Cf. Art. 4b AI-Act-EP-P; Art. 4b AI-Act-PA.

[74] Art. 14 (3) lit. a AI-Act-PA.

[75] Art. 14 (3) lit. b AI-Act-PA.

[76] Johann Laux, 'Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act' (2024) AI & Society 2853 (2858); Enqvist (n 60); Guillermo Lazcoz and Paul De Hert, 'Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-Requisites against Abdicating Responsibilities' [2022] Brussels Privacy Hub Working Paper, 10 <https://papers.ssrn.com/abstract=4016502> accessed 18 March 2024.

[77] Johann Laux, 'Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act' (2024) AI & Society 2853 (2858).

Provisional Agreement, which builds upon the EP's Proposal, contains a linguistically clearer provision.

Although the Proposals do not prescribe specific measures, they do specify certain minimum requirements that must be met for both built-in (Art. 14 (3) lit. a AI-Act-PA) and deployer-implemented (Art. 14 (3) lit. b AI-Act-PA) measures.

While the Commission Proposal required the user (now deployer) to *fully* understand the capacities and limitations of the high-risk AI-system[78], the EP Proposal allowed for just a *sufficient* understanding.[79] In this way, the EP proposal considers the fact that complete understanding would hardly or not at all be feasible in the case of (partially) autonomous systems, leaving room for more practicable solutions.[80] The Provisional Agreement rather follows the EP's position and calls for a *proper* understanding of relevant capacities and limitations of the high-risk AI-system.[81]

Both the Proposals and the Provisional Agreement leave open the form in which the AI-system's decision-making should be communicated to the human user. To ensure effective human oversight, a mere machine-readable explanation will probably not be sufficient. In fact, the Provisional Agreement seems to follow a technology- and innovation-friendly approach by acknowledging in Art. 14 (4) lit. c AI-Act-PA that different interpretation tools and methods are available and conceivable to meet the regulatory requirements. However, the field of interpretation methods and tools is highly dynamic.[82] Since the wording of Art. 14 (4) lit. c AI-Act-PA only refers to the *availability* of a certain method, it remains unclear if the method or tool must also be compliant with the current state of the art. However, it is difficult to imagine that the European legislator wanted to allow outdated technologies to suffice here. In any case, even outdated methods and tools might be covered by the wording if they continue to provide for *correct* interpretation results.

Furthermore, according to Art. 14 (4) lit. d AI-Act-PA, the human oversight measures must enable the natural persons, to whom human oversight is assigned, to be able to decide, in any particular situation, not to use the high-risk AI-system or otherwise

[78] Art. 14 (4) lit. a AI-Act-Com-P.

[79] Art. 14 (4) lit. a AI-Act-EP-P.

[80] Enqvist (n 62) 519.

[81] Art. 14 (4) lit. a AI-Act-PA.

[82] For an overview cf. below at III.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

disregard, override or reverse the output of the high-risk AI-system.[83] Recital 48 p. 3 AI-Act-PA specifies this by stating that such measures are intended to ensure, where appropriate, that the system is subject to built-in operational constraints that the system itself cannot override and that it is responsive to the human operator. This also counters the observation that certain AI-systems work to escape external dependencies to ensure their continued existence.[84] Ultimately, the system design must even provide for a stop/panic button according to Art. 14 (4) lit. e AI-Act-PA, which enables the natural person to interrupt the system's operation at any point in time. However, the final version limits this requirement to only bringing the AI-system to a standstill in a safe state.[85] This amendment, which has already been taken into account in the parliamentary version, is particularly useful for applications, where a more differentiated solution must be found.[86] This includes, among others, medical applications high-risk AI-systems.[87] For example, switching off a supporting AI-system (e.g. automated monitoring of vital signs during an operation) could otherwise have fatal consequences for patients.[88]

However, in the medical field of application, it is questionable whether the presumed superiority of humans over AI-systems applies when it comes to decisions that must be made with great precision or under time pressure.[89] Thus, application areas are conceivable in which human intervention would be neither possible nor desirable[90] and the unprecise wording of the commission text has been criticized accordingly.[91] Consequently, the EP also included the restriction in Art. 14 (4) lit. e AI-Act-EP-P, that the system must only come to a halt in a safe state, if the human interference

[83] Mazzini and Scalzo (n 40) 15f draw the comparison to Art. 22 GDPR and demand that the human oversight must be "meaningful rather than just a token gesture".

[84] Manuel Alfonseca and others, 'Superintelligence Cannot Be Contained: Lessons from Computability Theory' (2021) 70 Journal of Artificial Intelligence Research 65, 66.

[85] Art. 14 (4) lit. e AI-Act-PA.

[86] Cf. the criticism by Bomhard and Merkle (n 41) 281.

[87] Medical devices can be classified as high-risk AI-systems according to Art. 6 (1) lit. a, Annex I No. 11 AI-Act-PA.

[88] Ulrich M Gassner, 'Menschliche Aufsicht Über Intelligente Medizinprodukte Beitragsreihe KI' [2023] MPR 5, 9.

[89] Ulrich M Gassner, 'Menschliche Aufsicht Über Intelligente Medizinprodukte Beitragsreihe KI' [2023] MPR 5; Christian Straker and Maurice Niehoff, 'Die Regulierung Der Mensch-Maschine-Schnittstelle Algorithmischer Entscheidungssysteme' [2019] DSRITB 451.

[90] High-Level Expert Group on AI (n 8) para 65.

[91] Bomhard and Merkle (n 41) 281; Roos and Weitz (n 38) 847.

does not increase the risks or would negatively impact the performance in consideration of generally acknowledged state-of-the-art. However, this passage did not find its way into the final version, which is not understandable, at least in view of the examples mentioned above.

Finally, neither the Commission's nor the EP'S Proposal take on and develop ideas that have been previously included for example in the White Paper on AI. For example, it does not seem to be necessary that the output of the AI-system only becomes effective after it has been previously reviewed and validated by a natural person.[92] In addition, it is yet open whether real-time human oversight is necessary or dependent on the individual AI-system.[93] In this regard, the legislator has yet missed the chance to provide for "specific rules on frequency, scope or human involvement in certain sectors".[94]

ii. Criterion of Proportionality

The core element of EP's Proposal is the introduction of a proportionality criterion in Art. 14 (1) AI-Act-EP-P.[95] Accordingly, high-risk AI-systems shall be designed in such a way, that they can be effectively overseen by natural persons *as proportionate to the risks associated with those systems*. In addition, Art. 14 (3) AI-Act-EP-P provides that human oversight shall take into account the level of automation, and the context of the AI-system.[96] In previous text versions the consideration of the context of the AI-system could be derived from Art. 8 (2) AI-Act-Com-P. However, the other requirements were not included in the text, so they could at best be considered by interpreting them in accordance with fundamental rights.[97] The EP's Proposal, however, did not prevail the trialogue-negotiations in total. Instead, significant parts have been transferred from the prominent position in Section 1 to

[92] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5) 21.

[93] Bomhard and Merkle (n 41) 281; Initiative for applied Artificial Intelligence, 'Position Paper: Response to the 2020 European Commission's White Paper on AI' 8 <https://aai.frb.io/assets/files/Response-to-European-Commissions-White-Paper-on-AI-August-272020.pdf>.

[94] European Law Institute (n 2) 23.

[95] The proportionality criterion has already been proposed elsewhere, cf. for example European Law Institute (n 2) 22.

[96] Further criteria to consider according to Draft DIN EN ISO/IEC 22989:2023-04, p. 36 include the presence or absence of external monitoring, the degree of situated understanding of the system and the confidence with which the system can think and act in its environment, the degree of reactivity or responsiveness, the degree of adaptability to internal or external changes, requirements, or drives, the ability to assess one's own performance or suitability, including assessment against specified objectives, and the ability to decide and plan proactively in light of system goals, motivations, and drives.

[97] Similar with reference to the Council's Proposal

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

Section 3. Accordingly, Art. 14 (3) AI-Act-PA now provides that oversight measures shall be commensurate to the risks, level of autonomy and context of use of the AI-system. Additionally, oversight measures can be adopted as appropriate and proportionate to the circumstances.[98]

Despite the clear textual reference to some kind of proportionality criterion, the terms such as "appropriate", "proportionate", "context" or "circumstances" are not explained, but seem to stand independently. Nevertheless, there are some aspects to consider:

e) Risks Associated with and Context of the High-Risk AI-System

In the future, the proportionality test, which considers the anticipated risks to health, safety or fundamental rights[99] associated with the high-risk AI-system, will primarily have to evaluate the advantages and disadvantages of the AI application. Although this is, of course, a case-by-case consideration, the fact that the systems to be used will always be high-risk systems must be taken into account. Accordingly, high demands will have to be placed on human oversight and its intensity in general. Furthermore, the requirement to consider the anticipated risk is already known from other Regulations, such as the Commission Proposal for Art. 7 (2) lit. a Directive on Platform Work.[100] Accordingly, digital labor platforms shall evaluate the risks of automated monitoring and decision-making systems to the safety and health of platform workers, in particular as regards possible risks of work-related accidents, psychosocial and ergonomic risks.[101] This provision might serve as an example for other use case scenarios of high-risk AI-systems, where the associated risks have to be assessed.

f) Level of Automation of the High-Risk AI-System

To continue, the criterion of proportionality can also include the level of automation of the high-risk AI-system. Accordingly, the requirements for human oversight would have to increase as the degree of automation of the system increases. Despite the fact, that neither the Commission's nor the EP's Proposal nor the Provisional Agreement provide for a classification or list of potential degrees of automation of AI-systems,

[98] Art. 14 (4) AI-Act-PA; this wording could already be found in the Council's and EP's mandate.

[99] Art. 14 (2) AI-Act-PA.

[100] Proposal for a Directive of the European Parliament and of the Council on improving working conditions in platform work, COM(2021) 762 final.

[101] European Law Institute (n 2) 23.

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

sets of classification have been proposed (among others[102]) by the High Level Expert Group on AI and just recently also by the International Organization for Standardization. This criterion is not completely novel, but has already been mentioned before in the White Paper on AI.[103] As will be shown in the next subsection, the HLEG builds upon the approach mentioned in the White Paper and developed a more differentiated classification approach.[104]

g) High Level Expert Group on AI (HLEG)

The HLEG wants to conduct a differentiation of the degrees of automation based on the governance mechanisms in place. The group distinguishes between human-in-the-loop (HITL), human-on-the-loop (HOTL), and human-in-command (HIC) approaches. HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases would be neither possible nor desirable.[105] Human-on-the-loop refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. Human-in-command refers to the capability to oversee the overall activity of the AI-system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the AI-system in any particular situation.[106] In HIC applications, the human has the decision-making power not to use an AI-system in a particular situation, to allow a certain level of human discretion when using the system, or to ensure that a decision made by the system can be overridden. In addition, the person exercising oversight must also have the necessary authority to exercise oversight consistent with their particular assignment.[107]

---

[102] Johann Laux, 'Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act' (2024) AI & Society 2853 (2857) (with further reference) proposes a differentiation between first-degree (decisional output will depend on a human judgement of the AI's prediction) and second-degree human oversight (human involvement in terms of subsequent reviews and audits); Thomas B Sheridan and Raja Parasuraman, 'Human-Automation Interaction' (2005) 1 Reviews of Human Factors and Ergonomics 89; Victor Riley, 'A General Model of Mixed-Initiative Human-Machine Systems' (1989) 33 Proceedings of the Human Factors Society Annual Meeting 124; Raja Parasuraman and Victor Riley, 'Humans and Automation: Use, Misuse, Disuse, Abuse' (1997) 39 Human Factors: The Journal of the Human Factors and Ergonomics Society 230.

[103] White Paper on Artificial Intelligence - A European approach to excellence and trust (n 5).

[104] Lazcoz and De Hert (n 68) 10 et seqq.

[105] High-Level Expert Group on AI (n 8).

[106] High-Level Expert Group on AI (n 8).

[107] High-Level Expert Group on AI (n 8).

As already shown, the temporal scope of Art. 14 AI-Act-PA is not limited to the design level of the AI-system, but is comprehensive. Therefore, features and approaches which exclusively target the design stage are not adopted by the AI-Act-PA. Some features and approaches of the HLEG, however, can be found in Art. 14 (4) AI-Act-PA: For example, the ability to decide when and how to use the AI-system in any particular situation can be found in Art. 14 (4) lit. d AI-Act-PA. Additionally, the requirement that the person exercising oversight must also have the necessary authority is familiar from Art. 29 (1a) AI-Act-PA.

However, this classification of AI-systems, which differentiates according to only three categories, can only provide a first approach. Despite the fact that the individual categories also contain requirements that can be found in the Proposals, the large number and diversity of current and future AI-systems make a more precise differentiation necessary to determine proportionate human oversight measures.

h) International Organization for Standardization (ISO)

Most recently, ISO has developed a draft for a Technical Standardization of Artificial Intelligence in the Technical Committee ISO/IEC JTC 1, which has been adopted by the Technical Committee CEN/CLC/JTC 21 of the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENLEC). Despite the (democratic-theoretical) problems associated with the use of private norm-setting organizations[108] – in particular the missing participation of consumer organizations in private standardization processes[109] – their norms are also of particular importance for the field of artificial intelligence. This is due to Art. 40 AI-Act-PA, which provides for a presumption of conformity with the requirements set out in Title III Chapter 2 AI-Act-PA if high-risk AI-systems are in conformity with harmonized standards or parts thereof.[110] Thus, harmonized standards might not have a binding character for the addressees of the AI Act in the first place. However, it is likely that providers and deployers will follow this standard, because then they will not have to interpret the requirements of

[108] Cf. Martin Ebers, 'Ebers: Standardisierung Künstlicher Intelligenz und KI-Verordnungsvorschlag' [2021] Recht Digital 588, 593f with further references.

[109] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' [2021] Computer Law Review International 97, 105.

[110] An overview of contemporary standardization efforts is given in German Institute for Standardization and German Commission for Electrical, Electronic & Information Technologies of DIN and VDE, 'Position Paper on the EU "Artificial Intelligence Act"' (2021) <https://www.din.de/resource/blob/800324/c50ed443e81c47f8860b3f5c2b3b0742/21-06-din-dke-position-paper-artificial-intelligence-act-data.pdf>; Ebers (n 97).

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

the AI-Act on their own[111] and provide a justification why their measures are *equivalent* to the harmonized standards.[112] The CJEU also recently ruled that no fee may be charged for access to harmonized standards.[113] This means that the standards are now easier to access providing for an even greater incentive to use them.

Draft ISO/IEC 22989:2023-04 differentiates between six degrees of autonomy, automation, and heteronomy.[114] The classification ranges from Level 0 (No automation) to Level 6, where the system is capable of changing its intended area of application or goals without outside intervention, control or oversight. Below that, the system can either perform its entire task (Level 5, complete automation) or parts thereof (Level 4, high Level of automation) without external intervention. On Level 3 (Conditional/limited automation) an external agent is ready to step in when needed, while a durable and specific performance of a system is running independently. The last level of automation (Level 2) is referred to as partial automation, where some sub-functions of the system are fully automated, while the system is under the management of an external agent.

This standardization draft provides for a more sophisticated differentiation and it is therefore likely to be applied in practice. It can provide a sophisticated basis especially because providers and deployers have to take the level of autonomy of the AI-system into account when selecting particular oversight measures.[115] However, even this approach must consider that AI is an enabling technology subject to rapid change in research and development.[116] Therefore, the current update cycles of technical standards - at least where these standards relate to artificial intelligence - must be reviewed and, if necessary, adapted to the changed framework conditions. In addition, when classifying levels of automation, it is also important to consider whether a single AI-system can evolve independently during its use in such a way that it may reach a higher or lower level of automation. Consequently, regular and continuous re-evaluation of the proportionality of the specific human oversight measures must also be considered on the providers and deployers side.

---

[111] Veale and Zuiderveen Borgesius (n 88) 105.

[112] Art. 41 (4) AI-Act-PA.

[113] CJEU Case 588/21 P *Public.Resource.Org and Right to Know/ Commission a.o.* [2024] *to be published.*

[114] International Organization for Standardization, Draft ISO/IEC 22989:2023-04, 36.

[115] Art. 14 (3) AI-Act-PA.

[116] Ebers (n 31) 593.

Finally, it is also unclear whether one standard is sufficient for a large number of different high-risk AI-systems. After all, different fields of application can also entail different requirements, which must be reflected in the technical standards. Thus, a precise delineation of the scope of application of AI technical standards is also necessary.[117]

iii. AI Literacy

The success or failure of human oversight will ultimately be determined not only by the technical and organizational requirements, but also by the competence level of the natural person entrusted with human oversight.[118] This is because the best system can lead to wrong decisions if the natural person in charge of oversight does not or does not correctly interpret the output of the AI-system or is not aware of the risks involved in automated decision-making. In general, AI Literacy follows a two-dimensional approach: First, the human agent must have the competence to understand the subject matter. Second, he must have the ability to critically review and assess the AI-systems output.[119] The legislator now addresses these problems[120] that may arise in human handling of automated decisions.

First, the human oversight measures must enable the natural person entrusted with the human oversight to properly understand the capacities and limitations of the high-risk AI-system (Art. 14 (4) lit. a AI-Act-PA). As explained before, the EP advocated for a narrow understanding, limiting the necessary understanding to the relevant parameters.[121] In this respect, an appropriate compromise seems to have been found in the trialogue-negotiations. In order to fulfill this task successfully, the Commission proposed that the natural person should build up and continuously develop the corresponding and necessary competencies, conduct training, and have the authority

[117] Hadrien Pouget and Ranj Zuhdi, 'AI and Product Safety Standards Under the EU AI Act' <https://carnegieendowment.org/2024/03/05/ai-and-product-safety-standards-under-eu-ai-act?utm_source=substack&utm_medium=email> accessed 18 March 2024.

[118] A detailed oversight of possible problems and challenges can be found in Parasuraman and Riley (n 72) 234 ff.

[119] Kyriakos Kyriakou and Jahna Otterbacher, 'In humans, we trust' (2023) Discover Artificial Intelligence, Vol. 3 article number 44, Sec. 5.

[120] Cf. for an overview Jan Biermann, John J Horton and Johannes Walter, 'Algorithmic Advice as a Credence Good' <https://papers.ssrn.com/abstract=4326911> accessed 8 August 2023. Parasuraman and Riley (n 91). Andreas Fügener and others, 'Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI' (2021) 45 MIS Quarterly 1527.

[121] Art. 14 (4) lit. a AI-Act-EP-P.

to carry out human oversight.[122] The EP has partly taken up this proposal and transferred it to the enacting part of the Regulation, which will certainly provide for a greater impact. In particular, according to Art. 14 (1) s. 2, Art. 4b (2) AI-Act-EP-P, providers and developers of AI-systems shall take measures to ensure a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI-systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI-systems are to be used in, and considering the persons or groups of persons on which the AI-systems are to be used. In particular, this includes skills, knowledge and understanding that allows providers, users and affected persons to take into account their respective rights and obligations under the AI-Act.[123] Training and education should ensure, that these addressees are able to make an informed deployment of AI-systems, as well as gain awareness about the opportunities and risks of AI and possible harm it can cause.[124]

The Provisional Agreement is based on the EP's understanding going beyond the individual technical skill level. According to Art. 3 lit. bh AI-Act-PA[125], 'AI literacy' refers to skills, knowledge and understanding that allows providers, users (meant to be deployers) and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause.

Second, the natural person must be able to correctly interpret the high-risk AI-system's output, considering in particular the characteristics of the system and the interpretation tools and methods available.[126] As described before, which interpretation tools and methods available are to be used is to be assessed in the individual case and the measures must only provide for the *possibility* of correct interpretation.

Third, the natural person must remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI-system ('automation bias').[127] This should apply in particular to high-risk AI-systems that are

[122] Recital 48 s. 3 AI-Act-Com-P.

[123] Recital 9b AI-Act-EP-P.

[124] Recital 9b AI-Act-EP-P.

[125] The sections in Art. 3 AI-Act-PA do not yet appear to be subject to a consecutive enumeration system, so reference is meant to be made to the definition of "AI Literacy".

[126] Art. 14 (4) lit. c AI-Act-Com-P.

[127] Art. 14 (4) lit. b AI-Act-Com-P.

used to provide information or recommendations for decisions made by natural persons. Since HO measures can also be achieved with appropriate human-machine interface tools, a technical solution is also conceivable. This measure is particularly desirable, because studies show, that natural person may be willing to delegate high-stake decisions to AI, as they believe that AI performs better in this context.[128] In addition, however, suitable skilling and reselling programs should also be offered to ensure competent handling of the AI-system.

iv. Requirements for AI-systems for the Biometric Identification of Natural Persons

In the case of high-risk AI-systems intended to be used for the biometric identification of natural persons[129], the requirements referred to in Art. 14 (3) AI-Act-Com-P shall be such that, in addition, the user does not take any action or decision solely on the basis of the identification result produced by the AI-system until this has been verified and validated by at least two natural persons possessing the necessary competence, training and authority. Although this sets particularly high requirements, the scope of application is conceivably small. On the one hand, real-time biometric remote identification systems are only permitted at all in narrowly defined exceptional cases according to Art. 5 (1) lit. d, (2)-(4) AI-Act-Com-P. On the other hand, the Czech Presidency has recently proposed that Art. 14 (5) AI-Act-Com-P should not apply to measures relating to border control, in cases where Union or national law considers the application of Art. 14 (5) AI-Act-Com-P to be disproportionate.

v. Human oversight as one component in the system of requirements for high-risk AI-systems

Finally, it should be noted that human supervision is only one component of the requirements for high-risk AI-systems. In this regard, Art. 14 (2) AI-Act-PA already clarifies that human oversight shall aim at preventing or minimizing the risks to health, safety or fundamental rights, in particular when such risks persist notwithstanding the application of other requirements set out in Chapter II. This clarifies that the requirements of Chapter II are always cumulative and not alternative.

---

[128] Markus Christen and others, 'Who Is Controlling Whom? Reframing "Meaningful Human Control" of AI-systems in Security' (2023) 25 Ethics and Information Technology 5 with reference to Suzanne Tolmeijer and others, 'Capable but Amoral? Comparing AI and Human Expert Collaboration in Ethical Decision Making', *CHI Conference on Human Factors in Computing Systems* (ACM 2022) <https://dl.acm.org/doi/10.1145/3491102.3517732> accessed 10 August 2023.

[129] Annex III No. 1 lit. a AI-Act-Com-P; for a more detailed overview of the requirements of biometric real-time identification by law enforcement Agencies cf.

However, it is not clear from the Proposals whether (for example) lower requirements can be placed on human oversight measures if other requirements of Art. 9-15 AI-Act-PA are fulfilled on a basis that goes beyond the mandatory requirements. In any case, the margin should not be wide, since Art. 9-15 AI-Act-PA only address high-risk AI-systems anyway. Higher requirements for HO measures are therefore particularly conceivable if the other risk management systems are not or not sufficiently promising.[130] Similarly, the HLEG proposes a flexible approach, making the requirements for human oversight measures also dependent on the fulfilment of other requirements for trustworthy and ethical AI.[131]

III. Multidisciplinary View on the Requirements of Human Oversight

In the following, we analyze the AI-Act's requirements on human oversight in the light of Information Systems (IS) research and related domains. Considering the AI-system not as an alone standing technology, but as a socio-technical system is crucial to understand its potential effects not only on the individuum but on a society itself. AI's integration into information and decision systems is continuously striving. Therefore, becoming aware of the potentials but also of the negative (sometimes unintended) effects, not only from a regulatory perspective, is important for achieving trustworthy and ethical AI.

A. Nuances of Human Oversight

According to Declaration No. 9 lit. c, human oversight should consist of monitoring all outcomes that affect the safety and fundamental rights of people. Therefore, regulatory bodies consider human oversight as a mechanism to reduce potential risks stemming from AI-systems especially in high-risk scenarios. When looking into IS research, a more nuanced yet diverse picture is drawn. First, IS research defined multiple similar and sometimes overlapping research streams around human oversight. A human supervising an AI's behavior is also referred to as hybrid intelligence[132] or AI augmentation.[133] Both are situated in the broader research stream

[130] Enqvist (n 60) 517.

[131] High-Level Expert Group on AI (n 8) 16; Johanna Hahn, 'Die Regulierung Biometrischer Fernidentifizierung in Der Strafverfolgung Im KI-Verordnungsentwurf Der EU-Kommission' [2023] Zeitschrift für Digitalisierung und Recht 142.

[132] Dominik Dellermann and others, 'Hybrid Intelligence' (2019) 61 Business & Information Systems Engineering 637; Ece Kamar, 'Directions in Hybrid Intelligence: Complementing AI-systems with Human Intelligence.' (2016).

[133] Paul R Daugherty and H James Wilson, *Human+ Machine: Reimagining Work in the Age of AI* (Harvard Business Press 2018); Dirk Lindebaum, Mikko Vesa and Frank Den Hond, 'Insights from

of human-AI collaboration that is part of research on human-computer interaction. In parallel, IS scholars generally interpret human oversight not only as a risk reduction mechanism but rather as a form of improved decision making. The authors emphasize that amplifying, rather than replacing, human capabilities[134] by AI technology compensates for each other's limitations, resulting in superior performance than each could achieve independently. While humans bring social and contextual assessment skills to the table, AI is able of processing of huge data sets, objective decision making, and consistent and fast decision making. However, there is also research that criticizes placing humans in-the-loop. A study by Fügener et al. found that humans in a group assisted by AI can lose their individual knowledge by converging towards similar responses.[135] As the collective accuracy of the human group increases, the unique knowledge of each individual diminishes. This can lead to serious consequences in real life scenarios. As a potential mitigation strategy to this problem, the authors suggest deploying personalized AI advice.

B. Humans' Understanding of AI Reasoning

Current drafts propose the need to establish full understanding of the capacities and limitations of the high-risk AI-system (Art. 14 (4) lit. a AI-Act-Com-P), or at least a narrow understanding of the relevant parameters.[136] From a psychological perspective this can be quite challenging to achieve in reality. Not only are AI-systems characterized by continuous learning and evolve over time[137], but also their complexity of reasoning makes it practically impossible for a human user to understand and reconstruct all potential edge cases. In a time when AI-systems, especially those built on deep neural networks, become more advanced, the complexity of their reasoning surpasses what humans can easily understand. As a result, individuals find it difficult to predict the outcomes of decisions made by AI models[138] and struggle to assess an AI's quality and abilities. This challenge makes it

"the Machine Stops" to Better Understand Rational Assumptions in Algorithmic Decision Making and Its Implications for Organizations' (2020) 45 Academy of Management Review 247.

[134] Sebastian Raisch and Sebastian Krakowski, 'Artificial Intelligence and Management: The Automation–Augmentation Paradox' (2021) 46 Academy of management review 192.

[135] Fügener and others (n 108).

[136] Art. 14 (4) lit. a AI-Act-EP-P.

[137] Pär J Ågerfalk, 'Artificial Intelligence as Digital Agency' (2020) 29 European Journal of Information Systems 1.

[138] Alon Jacovi and others, 'Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI' (2021).

hard for humans to accurately evaluate the recommendations provided by AI.[139] A promising solution to this problem is XAI. While a separate research community is forming around XAI, expectations are growing that this technology will significantly support appropriate reliance on the AI's output.

C. XAI as a Tool to Support Human Oversight

Human oversight can only be carried out if humans can evaluate the quality of the AI output and correct any errors that occur. Information about the reasoning of the AI, i.e. how the AI arrived at a certain output, is an important prerequisite for human decision-making in order to fulfill this task. While the algorithms of some AI models can be understood by nature (white-box AI), modern, very powerful approaches in particular (e.g. deep neural networks or random decision trees) are considered non-transparent (black-box AI).

XAI is a technology that promises to shed light on the decision-making process of this black-box AI and thus make it comprehensible for a human user. To do so, XAI usually employs a transparent (white-box) algorithm to identify an understandable function that closely mirrors the outcomes of an opaque (black-box) algorithm.[140] Here, researcher differentiate between global and local explanations. *Global* explanations draw a broader picture on the AI's overall decision process and are therefore especially useful in the development and design stage of an AI-system. *Local* explanations, however, focus on revealing the causal relationships between an input and the corresponding specific output of an AI model.[141] Against this background, it's crucial to acknowledge that in attempts to replicate the reasoning of black-box AIs through white-box algorithms, the explanations provided by XAI can be imprecise or misleading. Consequently, users may encounter the issue of "ersatz understanding," a situation where individuals think they have a clearer insight into AI decisions through XAI data, even though such understanding doesn't align with the actual workings of the AI.[142] In parallel, recent research calls for caution when deploying XAI because it can have a negative effect on decision performance. For instance, in a study, Fuegener et al. discovered that, contrary to expectations, providing explanations about AI decisions actually resulted in lower accuracy in

[139] Jan Biermann, John J Horton and Johannes Walter, 'Algorithmic Advice as a Credence Good' [2022] ZEW-Centre for European Economic Research Discussion Paper.

[140] Boris Babic and others, 'Beware Explanations from AI in Health Care' (2021) 373 Science 284.

[141] Mengnan Du, Ninghao Liu and Xia Hu, 'Techniques for Interpretable Machine Learning' (2019) 63 Communications of the ACM 68.

[142] Babic and others (n 127).

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

human decision-making.[143] Similarly, in a separate experiment by Poursabzi-Sangdeh, it was initially hypothesized that participants who interacted with a transparent AI model would be better at spotting and correcting errors than those working with an opaque AI model.[144] However, the findings indicated the opposite effect.

IV. Conclusion

In conclusion, the legislative process represents a notable step forward in addressing the complexities of human oversight. By focusing on design, proportionality, and AI literacy, the legislator provided the groundwork for a comprehensive approach that ensures the responsible and effective use of high-risk AI-systems while accounting for the dynamic landscape of technological advancement and human capability. However, the paper has shown, that there is still leeway for the final text of the AI-Act. In this regard, the legislator should try to strike a careful balance between human intervention and the autonomy of AI-systems, particularly in critical domains like healthcare. In order to fulfil this task a critical engagement with the IS perspective is advisable. Understanding AI reasoning poses significant challenges due to the complexity and continual evolution of AI systems, especially those based on deep neural networks. Explainable AI (XAI) emerges as a solution to enhance human understanding of AI decisions. XAI aims to render opaque AI processes transparent through white-box algorithms, offering both global and local explanations. However, there are concerns about the accuracy and effectiveness of XAI explanations, leading to potential decreased decision performance in some cases.

[143] Fügener and others (n 108).

[144] Forough Poursabzi-Sangdeh and others, 'Manipulating and Measuring Model Interpretability' (2021).

University of Vienna Law Review, Vol. 9 No 3 (2025), pp. 1-34, https://doi.org/10.25365/vlr-2025-9-3-1.

V. Bibliography

Adadi A and Berrada M, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' (2018) 6 IEEE Access 52138

Ågerfalk PJ, 'Artificial Intelligence as Digital Agency' (2020) 29 European Journal of Information Systems 1

Alfonseca M and others, 'Superintelligence Cannot Be Contained: Lessons from Computability Theory' (2021) 70 Journal of Artificial Intelligence Research 65

Arrieta AB and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' <http://arxiv.org/abs/1910.10045> accessed 10 August 2023

Babic B and others, 'Beware Explanations from AI in Health Care' (2021) 373 Science 284

Becker D and Feuerstack D, 'Der Neue Entwurf Des EU-Parlaments Für Eine KI-Verordnung' [2024] Multimedia und Recht 22

Biermann J, Horton JJ and Walter J, 'Algorithmic Advice as a Credence Good' <https://papers.ssrn.com/abstract=4326911> accessed 8 August 2023

Biermann J, Horton JJ and Walter J, 'Algorithmic Advice as a Credence Good' [2022] ZEW-Centre for European Economic Research Discussion Paper

Bomhard D and Merkle M, 'Europäische KI-Verordnung' [2021] Recht Digital 276

Christen M and others, 'Who Is Controlling Whom? Reframing "Meaningful Human Control" of AI-systems in Security' (2023) 25 Ethics and Information Technology 10

Cocito C and Hert PD, 'The Use of Declarations by the European Commission: "Careful with That Axe, Eugene". — The Digital Constitutionalist' (14 March 2023) <https://digi-con.org/the-use-of-declarations-by-the-european-commission-careful-with-that-axe-eugene/> accessed 18 March 2024

Daugherty PR and Wilson HJ, *Human+ Machine: Reimagining Work in the Age of AI* (Harvard Business Press 2018)

Dellermann D and others, 'Hybrid Intelligence' (2019) 61 Business & Information Systems Engineering 637

Du M, Liu N and Hu X, 'Techniques for Interpretable Machine Learning' (2019) 63 Communications of the ACM 68

Ebers M, 'Ebers: Standardisierung Künstlicher Intelligenz Und KI-

Verordnungsvorschlag' [2021] Recht Digital 588

Enqvist L, '"Human Oversight" in the EU Artificial Intelligence Act: What, When and by Whom?' (2023) 15 Law, Innovation and Technology 508

'European Declaration on Digital Rights and Principles | Shaping Europe's Digital Future' (7 February 2023) <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles> accessed 9 August 2023

European Law Institute, 'Guiding Principles for Automated Decision-Maing in the EU' <https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Innovation_Paper_on_Guiding_Principles_for_ADM_in_the_EU.pdf> accessed 10 August 2023

Fügener A and others, 'Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI' (2021) 45 MIS Quarterly 1527

Gassner UM, 'Menschliche Aufsicht Über Intelligente Medizinprodukte Beitragsreihe KI' [2023] MPR 5

German Institute for Standardization and German Commission for Electrical, Electronic & Information Technologies of DIN and VDE, 'Position Paper on the EU "Artificial Intelligence Act"' (2021) <https://www.din.de/resource/blob/800324/c50ed443e81c47f8860b3f5c2b3b0742/21-06-din-dke-position-paper-artificial-intelligence-act-data.pdf>

Hahn J, 'Die Regulierung Biometrischer Fernidentifizierung in Der Strafverfolgung Im KI-Verordnungsentwurf Der EU-Kommission' [2023] Zeitschrift für Digitalisierung und Recht 142

Heil M, 'Die Neue KI-Verordnung (E) – Regulatorische Herausforderungen Für KI-Basierte Medizinprodukte-Software' [2022] Zeitschrift für das gesamte Medizinproduktrecht 1

High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI' <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Initiative for applied Artificial Intelligence, 'Position Paper: Response to the 2020 European Commission's White Paper on AI' <https://aai.frb.io/assets/files/Response-to-European-Commissions-White-Paper-on-AI-August-272020.pdf>

Jacovi A and others, 'Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI' (2021)

Kamar E, 'Directions in Hybrid Intelligence: Complementing AI-systems with Human Intelligence.' (2016)

Lazcoz G and De Hert P, 'Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-Requisites against Abdicating Responsibilities' [2022] Brussels Privacy Hub Working Paper <https://papers.ssrn.com/abstract=4016502> accessed 18 March 2024

Lindebaum D, Vesa M and Den Hond F, 'Insights from "the Machine Stops" to Better Understand Rational Assumptions in Algorithmic Decision Making and Its Implications for Organizations' (2020) 45 Academy of Management Review 247

Mazzini G and Scalzo S, 'The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts' [2022] SSRN Electronic Journal <https://www.ssrn.com/abstract=4098809> accessed 11 August 2023

Parasuraman R and Riley V, 'Humans and Automation: Use, Misuse, Disuse, Abuse' (1997) 39 Human Factors: The Journal of the Human Factors and Ergonomics Society 230

Pouget H and Zuhdi R, 'AI and Product Safety Standards Under the EU AI Act' <https://carnegieendowment.org/2024/03/05/ai-and-product-safety-standards-under-eu-ai-act?utm_source=substack&utm_medium=email> accessed 18 March 2024

Poursabzi-Sangdeh F and others, 'Manipulating and Measuring Model Interpretability' (2021)

Rai A, 'Explainable AI: From Black Box to Glass Box' (2020) 48 Journal of the Academy of Marketing Science 137

Raisch S and Krakowski S, 'Artificial Intelligence and Management: The Automation–Augmentation Paradox' (2021) 46 Academy of management review 192

Riley V, 'A General Model of Mixed-Initiative Human-Machine Systems' (1989) 33 Proceedings of the Human Factors Society Annual Meeting 124

Roos P and Weitz CA, 'Hochrisiko-KI-Systeme Im Kommissionsentwurf Für Eine KI-Verordnung' [2021] Multimedia und Recht 844

Schucht C, 'Die Neue Architektur Im Europäischen Produktsicherheitsrecht Nach New Legislative Framework Und Alignment Package' [2014] Europäische Zeitschrift für Wirtschaftsrecht 848

Schwalbe G and Finzel B, 'A Comprehensive Taxonomy for Explainable Artificial Intelligence: A Systematic Survey of Surveys on Methods and Concepts' [2023] Data Mining and Knowledge Discovery <http://arxiv.org/abs/2105.07190> accessed 10

August 2023

Sheridan TB and Parasuraman R, 'Human-Automation Interaction' (2005) 1 Reviews of Human Factors and Ergonomics 89

Spindler G, 'Der Vorschlag der EU-Kommission für eine Verordnung zur Regulierung der Künstlichen Intelligenz (KI-VO-E)' [2021] Computer und Recht 361

Straker C and Niehoff M, 'Die Regulierung Der Mensch-Maschine-Schnittstelle Algorithmischer Entscheidungssysteme' [2019] DSRITB 451

Tolmeijer S and others, 'Capable but Amoral? Comparing AI and Human Expert Collaboration in Ethical Decision Making', *CHI Conference on Human Factors in Computing Systems* (ACM 2022) <https://dl.acm.org/doi/10.1145/3491102.3517732> accessed 10 August 2023

Veale M and Zuiderveen Borgesius F, 'Demystifying the Draft EU Artificial Intelligence Act' [2021] Computer Law Review International 97

Wiebe G, 'Produktsicherheitsrechtliche Betrachtung Des Vorschlags Für Eine KI-Verordnung' [2022] Betriebs-Berater 899

Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Fostering a European approach to Artificial Intelligence [COM(2021) 205 final]

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence for Europe 2018 [COM(2018) 237 final]

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Coordinated Plan on Artificial Intelligence 2018 [COM/2018/795 final]

White Paper on Artificial Intelligence - A European approach to excellence and trust 2020