Accountability

WHITE PAPER

Trustworthy artificial intelligence

24 January 2024

Kerstin Waxnegger*, Sebastian Scher*, Simone Kopeinik*, Tomislav Nad**, and Dominik Kowald*

* Know-Center GmbH

** SGS Digital Trusts Services GmbH

Partners of SGS

Accountability

Contents:

Introduction

As technology advances, artificial intelligence (AI) is taking over more and more tasks seemingly independently. But the use of AI also poses risks. For example, an intelligent algorithm might discriminate against people and intelligent robots (including self-driving cars) can cause major harm. Thus, the question of who is responsible in the event of AI-related damages or accidents arises and legislators as well as other stakeholders all around the world are engaged in proposing and creating regulatory frameworks and guidelines for AI.

This white paper approaches the question of addressing legal accountability and provides an overview of the most important AI regulations and regulatory initiatives as well as AI certification. A comparison of different regulations, among them the EU AI Act, shall create a better understanding of the field. Additionally, open challenges and issues are tackled. Eventually, we conclude with a short summary.

Definition of terms

There is no standardised definition of "trustworthy AI". According to the EU's High-Level-Expert-Group on AI, AI can be perceived as "trustworthy", when it is "developed, deployed, and used in ways that not only ensure its compliance with all relevant laws and its robustness but especially its adherence to general ethical principles." [TLS21; AIHLEG19]

According to NIST, trustworthy AI includes these "essential building blocks": Validity and Reliability / Safety / Security and Resiliency / Accountability and Transparency / Explainability and Interpretability / Privacy / Fairness with Mitigation of Harmful Bias. [NIST]

Trustworthy AI is also often described as "responsible", "ethical" or "human-centered" AI. [BGT23]

As we can see, accountability plays a key role in trustworthy AI. Accountability generally refers to the responsibility and answerability for one's actions or decisions. When it comes to AI, accountability entails understanding who is responsible for the actions and outcomes of AI systems, as well as ensuring that appropriate mechanisms are in place to address any negative consequences or errors. In other words, an AI system will only be considered as trustworthy if someone is accountable for its errors.

Who can be held liable?

To create trust in AI, we need to develop technical and social robust AI systems that are compliant with applicable laws and guarantee compliance with ethical values [OMR+22]. AI regulation establishes specific rules to ensure legal certainty and address liability in a socially accepted way.

Different stakeholders can be held liable. The entity or individuals responsible for developing or manufacturing an AI system may be held liable if the harm is caused by a defect or negligence in the design, development or production of the AI technology. For example, the EU AI Act sets clear rules for developers/manufactures in the European Union (EU). The liability of developers might increase as they are the ones who affect the way AI systems act the most.

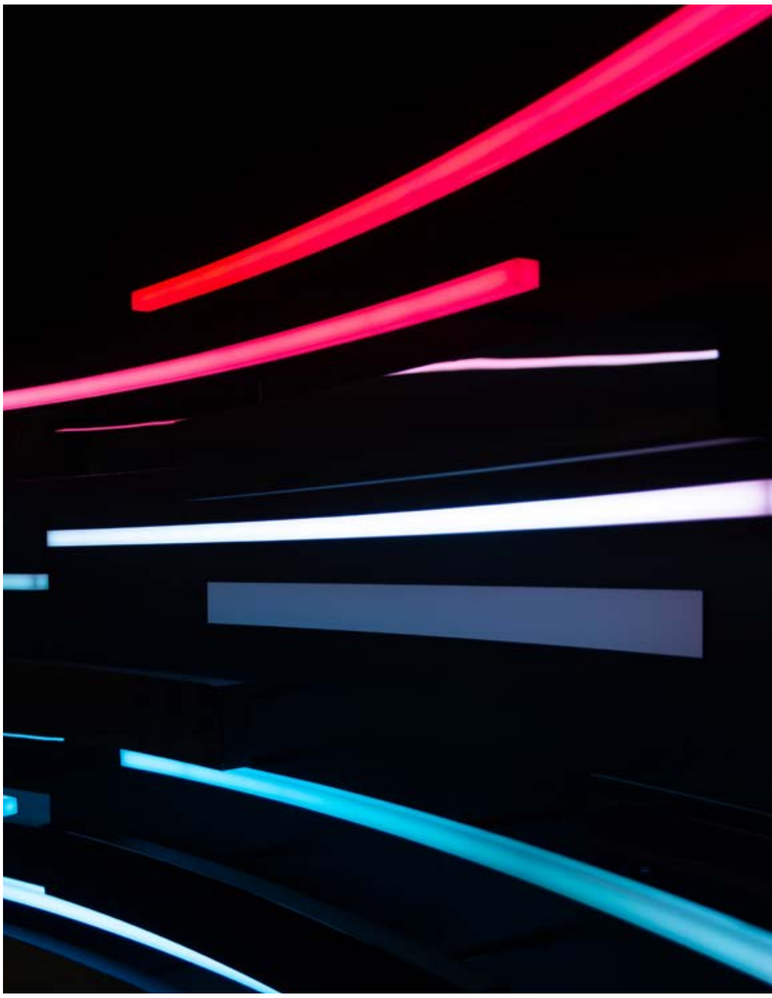
If the AI system is operated or used negligently or inappropriately, the individual or organization operating or utilizing the AI system might also bear liability. This could include factors such as inadequate training, improper use or failure to implement appropriate safety measures. Yet, the more autonomous AI systems act, the less the operator can interfere and therefore be liable.

In cases where the AI system's training data is flawed, biased or contains inaccurate information, the entity or individuals responsible for providing the data may share liability if the harm is a result of the flawed data. The importance of providing correct data will increase, as AI can only act properly if it uses the right training data. For example, the EU Data Act shall ensure fairness by setting rules for the use of data generated by Internet of Things (IoT) devices.

Finally, if a regulatory framework is in place, failure to comply with applicable regulations or standards related to AI deployment may result in liability for certifying parties.

It is important to note that liability laws surrounding AI are still evolving, and the specific legal framework and liability allocation may differ depending on the jurisdiction and the circumstances of the case. As AI technology continues to advance, legal systems are adapting to address the unique challenges and complexities posed by AI-related harms. Within the last few years, there has been a notable evolution in the AI governance landscape, marked by governments putting forth policies to regulate AI technologies within their respective jurisdictions. These frameworks shall promote a beneficial use of AI and manage risks of AI by addressing liability. Therefore, regulations must be achievable and accessible. In other words: "Workable pragmatical outcomes" shall be achieved. [STI21]

As **AI** technology continues to advance, legal systems are adapting to address the unique challenges and complexities posed by AI-related harms.



Ways of regulating AI: current state and open issues

There are different ways of regulating AI. On one hand, new frameworks, like the EU AI Act, can be enacted. On the other hand, existing laws might be adapted or might already be applicable to AI. For example, the General Data Protection Regulation (GDPR) also refers to data breaches involving AI, and criminal law (which is neutral with respect to specific technologies) can deal with offences committed by using AI.

Whenever new rules for AI are adopted, we can divide between "bottom-up" and "top-down" approaches. A bottom-up approach means that a body, like an international organization, develops standards which can be used as basis for establishing new rules by legislators if needed. This approach bears the risk that regulations might be set into force after an incident has happened. In contrast, top-down approaches follow a more futuristic approach: The legislator sets clear standards which may even be in force before technological improvements are practically applicable. The best example for this approach is the EU AI Act.

Regulations might also be concepted as "soft law" or "hard law". Soft law includes non-binding recommendations, resolutions, guidelines and standards, like the OECD recommendation on AI, ISO/DIN/CEN-standards etc. While soft law is relatively easy to adopt, it lacks enforcement. In contrast, hard law, like the EU AI Act, includes binding rules and therefore ensures enforcement and legal security. Its disadvantages are that it might not be unable to keep up with rapidly changing technology, provisions might not align on a global level and jurisdictions are limited in scope.

Today, numerous organizations, national authorities and other stakeholders take an interest in regulating AI. Among them, for example, the Council of Europe, the European Union (EU), the OECD, the United Nations (UN), national governments, numerous non-governmental organizations (NGOs) and standardization and research organizations. Furthermore, international and interdisciplinary platforms on the topic were founded.

The EU AI Act[1]

The EU AI Act was proposed by the European Commission in 2021 and aims to be the "world's first comprehensive AI law" [EUP23]. It follows a horizontal, technology and risk-based approach. The aim is to address the use of AI systems in certain areas of application and the Act might be adopted according to technical and market developments when in force. Technical requirements shall be addressed via separate, harmonized European standards (CEN/CENELEC), which will together form the new legislative framework for AI in the EU. These standards are yet to be created.

The EU AI Act divides AI applications into different risk levels, ranging from "unacceptable risk" to "high risk", "limited risk" and "minimal risk". Each of these risk levels in turn provides for different consequences, such as a ban on use or specific transparency obligations. Social scoring systems, for example, are completely prohibited. High risk systems that pose a significant threat to the environment, security, human health and fundamental rights are subject to strict transparency and monitoring obligations and require conformity and impact assessments as well as a registration in an EU database. These systems shall include, for example, AI systems for education and vocational training, employment and workers management or systems for creditworthiness evaluation and law enforcement. AI systems that pose little or no risk, on the other hand, will remain largely unregulated and subject to voluntary standards, among these, for example, spam filters or AI systems used in video games.

In addition, the EU AI Act shall also contain specific regulations and a two-tiered approach for General Purpose AI (GPAI), which can be used for a variety of purposes. Generally, GPAI will require technical documentation, complying with EU copyright law and disseminating summaries of training material. If a GPAI poses a "systemic risk", additional requirements, among them conducting model evaluations and risk assessments, conducting adversarial testing, reporting serious incidents to the European Commission, ensuring cybersecurity and reporting on energy efficiency, must be met.

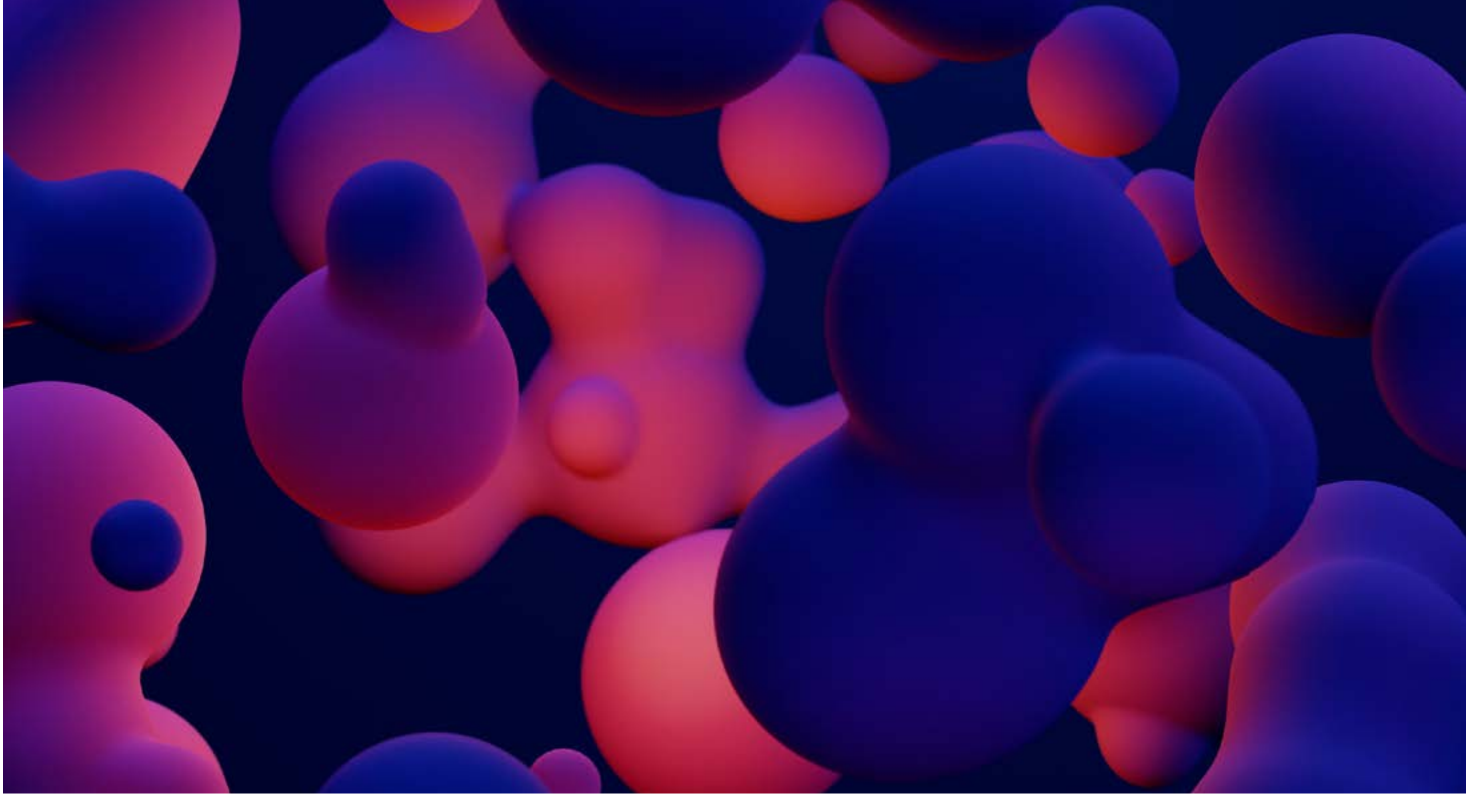
Military, defence and non-professional use; research, development and prototyping activities preceding the release on the market are not covered by the EU AI Act. Additionally, there are exemptions for research activities and provisions for AI under open-source licences. AI real-world laboratories will allow for a preliminary review of AI systems and there will be provisions for testing AI systems under real-life conditions. Finally, there should be an effective complaints procedure for citizens.

Fines for non-compliance range from 35 million euros or 7% of global turnover to 7.5 million euros or 1.5% of turnover, with caps for small and medium enterprises (SMEs) and start-ups.

The EU AI Act is surrounded by further EU provisions, among them for example the proposed AI liability directive, the Data Act, the Data Governance Act, the Digital Services Act, the Digital Markets Act, the GDPR, the Cybersecurity Act, the Cyber-Resilience Act, the NIS-RL, the product liability directive and other sectoral provisions.

> The aim of the AI Act is to address the use of **AI** systems in certain areas of application and might be adopted according to technical and market developments when in force.

[1] **Information regarding the final version of the AI Act is derived from information currently made public. See https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683.*



Comparison with other AI regulations

Understanding the similarities between (proposed) AI regulations, guidelines and frameworks is essential for understanding the AI regulation landscape and for building an interoperability between AI regulations on a global level. Therefore, we compare approaches for regulating AI from EU, US, Canada, UK and China as of December 2023. Please note again that the AI framework landscape is currently evolving and can change rapidly.

In the following, we summarize our present-day findings in a short overview. As previously described, the EU follows a hard law, horizontal, risk-based approach and uses a flexible definition on AI.

The EU AI Act also includes provisions for GPAI and is accomplished by further other hard law legislation (like the DSA, DMA etc.) as well as European standards. There are fines for non-compliance. Canada chose a similar approach when putting forth a horizontal, hard law approach in their proposed Artificial Intelligence and Data Act[2].

In contrast, the US takes a more contextual, decentralized, sectoral, soft law approach with light touch options and voluntary provisions. For instance, they introduced their "Blueprint for an AI Bill of Rights"[3]. In October 2023, the president of the US issued an Executive Order on Safe, Secure and Trustworthy Artificial Intelligence, which shall complement the Blueprint for an AI Bill of Rights and describes guiding principles and priorities when developing and using AI[4].

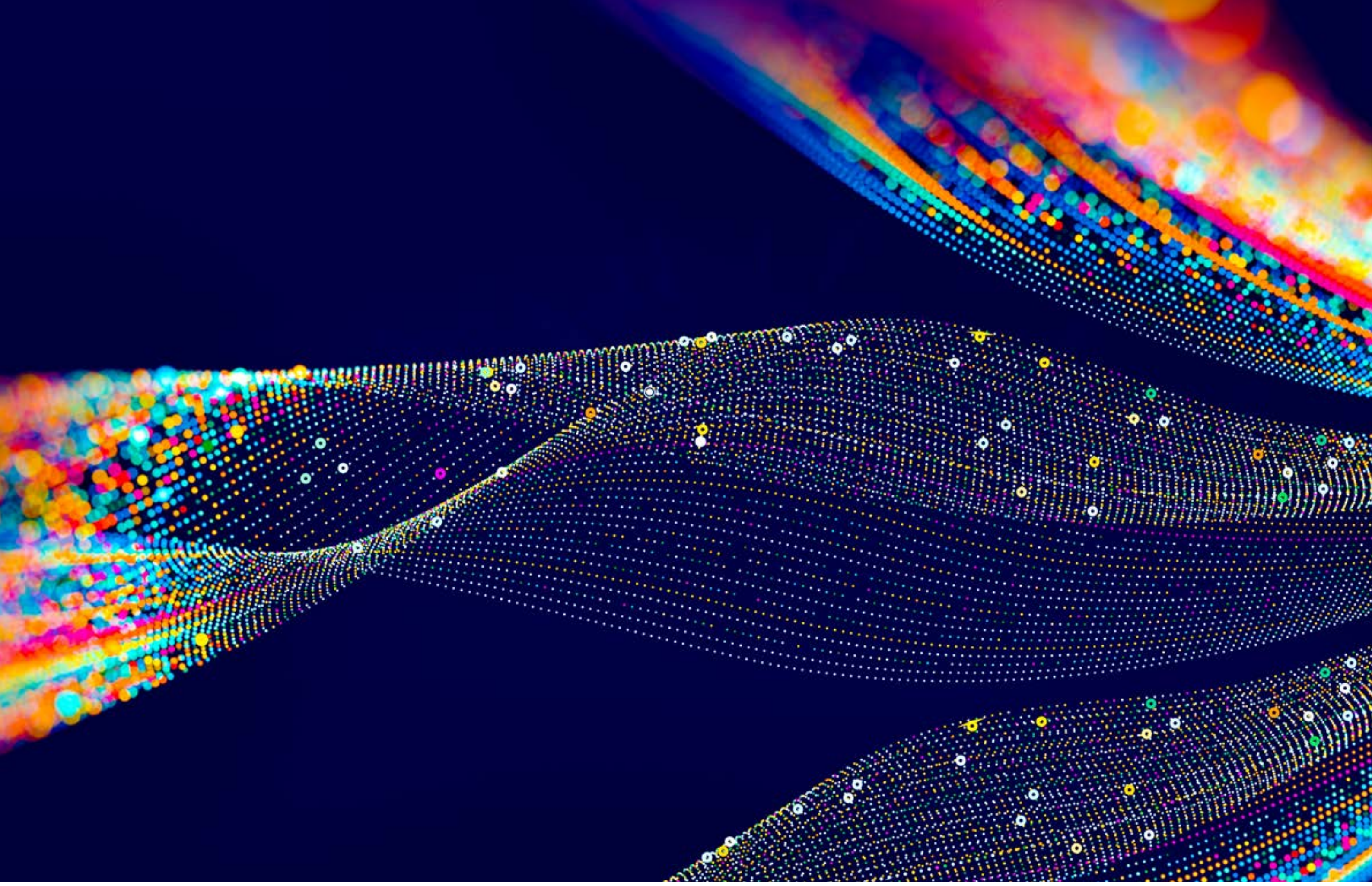
The UK also seeks a sectoral and context-based soft law approach, seemingly in favour of applying pre-existing rules t o AI. The UK also launched "AI Safety Institute" that shall task testing the safety of emerging types of AI[5]. China takes the middle part, implementing both soft law ethical principles for AI use and hard law provisions for specific technologies, ensuring enforcement through a national body. Additionally, China recently released new rules for generative AI[6].

As we can see, most frameworks follow a risk-based approach. They endorse similar key principles (accuracy and robustness, safety, non-discrimination, security, transparency and accountability, explainability and interpretability, data privacy) as well as the role of international standards. On one hand, UK/US frameworks are more flexible as soft law can be adopted easily. Yet, they lack enforcement. The proposed EU and Canadian regulations on the other hand, are more clarified and therefore easier to enforce [AOO22].

To summarize, there is no unified global approach to AI regulation. Therefore, AI frameworks might conflict with each other. Some frameworks and standards are also highly general and generic. They do not outline how they can be transferred into practice [TLS21]. A fragmented legal landscape leads to a lack of interoperability and a lack of enforcement, especially across national borders [ENG23]. It is important to align policies on a global level. Otherwise, frameworks lack meaningful enforcement and cannot assure that AI is responsible and trustworthy. For instance the Trade and Technology Council (TTC) aims to align EU and US frameworks via:

- Discussing measurement and evaluation of trustworthy AI
- Collaborating on AI technologies designed to protect privacy, and
- Jointly producing an economic study of AI's impact on the workforce [EUC23]

[2] *https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document.*
[3] *https://www.whitehouse.gov/ostp/ai-bill-of-rights/.*
[4] *https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.*
[5] *https://www.gov.uk/government/news/prime-minister-launches-new-ai-safety-institute.*
[6] *https://oecd.ai/en/dashboards/countries/China.*



3 Towards AI certification

Alongside the regulation of AI, certification of AI systems is becoming increasingly important. AI certification can indicate the current minimum performance standards developers must adhere to act compliant. They can also facilitate international cooperation and harmonization of AI regulations.

Certification means that compliance with rules/standards is checked by a third party (mainly an accredited body) – whether mandatory or non-binding. These checks shall help developers and users of AI to be able to tell if AI acts in a way that respects human rights if basic principles for trustworthy AI are met. Certification can enhance trust and provides legal security for developers and users. As already mentioned, complying with standards can also bring competitive advantages and international standards increase interoperability. But there are also downsides: strict rules and high costs might inhibit development and lead to distortion of competition as SMEs cannot afford certification. Standards might still be inconsistent on a world-wide level. Practicability might be low in the beginning and criteria might be difficult to verify. There are already standards for safety-critical applications which might be technology-neutral (e.g. medicine, aviation) and thus can be transferred to AI.

4 Summary

As AI technology evolves, more and more legal frameworks for AI are being put forth by various actors, like international organizations, national governments and the civil society. Hardly any binding regulations have been set into force so far. Yet, the proposed EU AI Act could be ground-breaking in terms of regulating AI.

Standards that define accountable use of AI and partly form the basis for AI regulations are still in development. These regulations vary on a world-wide level and might lead to a fragmented AI-regulatory landscape. It is important to align these standards on a global level to ensure interoperability and enforcement.

5 Acknowledgement

Know-Center is a leading European research center for big data, artificial intelligence (AI) and data-driven business models. Know-Center is a COMET center within COMET – Competence Centers for Excellent Technologies. This program is funded by the Austrian Federal Ministries for Climate Policy, Environment, Energy, Mobility, Innovation and Technology (BMK) and for Labor and Economy (BMAW), represented by Österreichische Forschungsförderungsgesellschaft mbH (FFG), Steirische Wirtschaftsförderungsgesellschaft mbH (SFG) and the Province of Styria, Wirtschaftsagentur Vienna and Standortagentur Tyrol GmbH.

References

[AIHLEG19] High-Level Expert Group on A 2019. Ethics Guidelines For Trustworthy Ai. *https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai*

[AOO22] Akinola, 2022. Comparative analysis regulatory of AI and algorithm in UK, EU and USA. *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4212588*

[BGT23] Brumen, B., Göllner, S., Tropmann-Frick, M., 2023. Aspects and Views on Responsible Artificial Intelligence, in: Nicosia, G., Ojha, V., La Malfa, E., La Malfa, G., Pardalos, P., Di Fatta, G., Giuffrida, G., Umeton, R. (Eds.), Machine Learning, Optimization, and Data Science. Springer Nature Switzerland, Cham, pp. 384–398

[ENG23] Engler, A., 2023. The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment. *https://www.brookings.edu/research/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/*

[EUC23] European Commission, 2023. EU-US Trade and Technology Council. *https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/stronger-europe-world/eu-us-trade-and-technology-council_en*

[EUP23] European Parliament, 2023. EU AI Act: first regulation on artificial intelligence. *https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence*

[NIST] National Institute of Standards and Technology, US Department of Commerce. *https://www.nist.gov/trustworthy-and-responsible-ai*

[OMR+22] Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., Agreda, S.G., 2022. To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. Technological Forecasting and Social Change 181. *https://doi.org/10.1016/j.techfore.2022.121763*

[STI21] Stix, C., 2021. Actionable Principles for Artificial Intelligence Policy: Three Pathways. Science and Engineering Ethics 27. *https://doi.org/10.1007/s11948-020-00277-3*

[TLS21] Thiebes, S., Lins, S., Sunyaev, A., 2021. Trustworthy artificial intelligence. Electronic Markets 31, 447–464. *https://doi.org/10.1007/s12525-020-00441-4*

Accountability

SGS.COM/DIGITAL

CONTACT US

SGS
Emerging Technology
Enquiry.Emerging-Technology@sgs.com

KNOW CENTER
Leading Research and Innovation Center for Trustworthy AI
https://know-center.at/
info@know-center.at

WHEN YOU NEED TO BE SURE