AI & Partners

Amsterdam - London - Singapore

EU AI Act

Artificial Intelligence Washing: Compliance and Awareness

July 2024

Amsterdam - London - Singapore

AI & Partners defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots professional services, regulatory interventions, and participating in industry groups such as AI Commons, we fight for fundamental rights in the artificial intelligence age.

This report was prepared by Sean Donald John Musch and Michael Charles Borrelli. For more information visit https://www.ai-and-partners.com/.

Contact: Michael Charles Borrelli | Chief Operating Officer | m.borrelli@ai-and-partners.com.

This report is an AI & Partners publication.

Our report identifies that the nascent phenomenon of AI washing, including its most relevant types and occurrences, is more likely to be understood by participants in the banking sector, such as investment firms and payment service providers. Additionally, AI risks can be addressed, either wholly or partially, through the EU AI Act as a robust regulatory framework given its heavy emphasis on both consumer and investor protection, paving the way to a safe, secure and trustworthy AI economy

About this report

This report is based on market research, publicly available data, and interviews with AI specialists in AI & Partners, financial services organisations, and relevant third-parties. Moreover, quotations provided on specific topics reflect those of AI specialists at AI & Partners to be as representative as possible of real-world conditions. All references to EU AI Act reflect the version of text valid as at 13 June 2024. Accessible here.

Contents

Executive Summary

Acknowleged by many as a hallmark milestone in artificial intelligence ("AI") regulation, the entry into force of the European Union ("EU") AI Act (the "EU AI Act") opens a new chapter on the road towards trustworthy AI. The regulation aims at empowering individuals by protecting their fundamental rights in an AI era and harmonising the patchwork of national data legislation across the EU to lay the foundation for a thriving digital market.

The regulation introduces certain rights for individuals and new obligations for providers, such as fundamental rights impact assessments ("FRIAs") and AI system breach notifications. Providers of AI systems, whether they are financial services institutions (e.g. investment firms, payment service providers etc.) or standalone technology providers, need a clear plan and strategy to best analyse the emerging phenomenon of AI washing, its impact, and how the key foundations to mitigate AI washing risk in the EU AI Act framework.

This white paper (the "Report") examines opportunities as to how EU AI Act compliance can address the risks posed by AI washing, prepare for a trustworthy AI economy, and substantially reduce future regulatory adjustment costs, in regards to AI washing, in the light of the existing global AI legislation. Moreover, it discusses the challenges faced by AI washing[1] and proposes a clear approach as to how embracing the EU AI Act can act as a counter-balancing force.

The Report takes a deeper look on potential AI Washing occurrences, based on those reported as reported by the Competent Authorities ("CAs") for its equivalent, 'Greenwashing', and provides suggestions on the adverse impact that AI Washing can have on the risks of providers and deployers and on consumers. The views of AI & Partners' on the materiality of AI Washing risk to firmss remain overall stable, yet it is becoming a more relevant issue. Reputational and operational risks are still considered as the potential types of risks most impacted by AI Washing, which is in line with the observation that litigation risk resulting from AI Washing is likely to be a rising trend in the next three years.

As firms expand their uptake and use of trustworthy AI products and are adapting their business models to meet challenges in relation to the transition towards a more AI-driven economy, addressing ai Washing is key to provide confidence in the market and maintain the trust of investors and consumers. This is relevant both in the context of specific products and services and for entity-level claims and commitments. In this context, market's best practices, regulation and supervision have a role to play in addressing integrity concerns.

Firms should take all necessary steps to ensure that trustworthy information provided is fair, clear, and not misleading. This includes observing key principles for trustworthy claims to be accurate, substantiated, up to date, fairly representative of the firm's overall profile or the profile of the product, and presented in an understandable manner. Firms should review and adapt their governance arrangements and internal processes to build safeguards against AI Washing, take a proactive approach in addressing data challenges, and consider the extent to which external verification and alignment with market guidance would support credibility of trustworthy AI products and/or targets.

[1] This term includes a practice whereby trustworthy-related statements, declarations, actions, or communications do not clearly and fairly reflect the underlying trustworthy profile of an AI solution, service, platform, product or other asset. This practice may be misleading to consumers, investors, or other market participants.

At the entity level, firms should substantiate forward-looking trustworthy commitments such as explainability and transparency claims with credible plans and strategies, provide clear and granular information on their trustworthy finance targets, and integrate AI Washing-related risks as part of their management of conduct, operational and reputational risks. At the product level, firms should establish and report clear criteria, definitions and indicators for products and/or services labelled as trustworthy. They should also apply rigor and closely engage with counterparties in designing trustworthy-linked products, in particular generative AI ("GenAI").

At a legislative and regulatory level, AI & Partners considers that the most effective way forward to address AI Washing by EU market participants is to focus on the finalisation and implementation of the existing and planned initiatives. The EU AI Act provides a key foundation to address several aspects of AI Washing concerns in the banking sector. This includes rules on consumer/investor protection that provide the legal basis for promoting the uptake of trustworthy AI, and trustworthy AI-related developments, including requiring disclosures in regards to general purpose AI ("GPAI") that should enhance transparency on trustworthy practices. However, the majority of these measures are still in the early stages of implementation, suggesting that benefits of these frameworks are not fully visible yet.

In the short term, priority should be given to supporting a robust implementation of the EU AI Act and to overcome the likely challenges. Efforts to address data, usability, consistency and international interoperability issues should be pursued.

AI & Partners recommends that firms pursue their planned and ongoing efforts and activities to identify and monitor AI Washing within the remit of their respective regulatory, legal and compliance obligations. Firms are encouraged to make use of the EU AI Act to the largest extent possible. Furthermore, building-up the capacities and fostering AI education are crucial to appropriately account for and mitigate AI risks.

Complementary to the recommendations to firms, AI & Partners continues to provide regulatory guidance on how to address AI Washing-related aspects under the EU AI Act and will facilitate knowledge sharing between firms, regulators and other industry stakeholders on the best market practices. Finally, AI & Partners will continue monitoring AI-Washing related trends and risks in the EU banking sector.

The Key Takeaways

- The EU AI Act represents a key regulatory framework in order to promote the uptake of responsible AI.
- AI Washing is a new phenomenon that gives the illusion of a product's trustworthiness, which has the potential mislead investors, consumers and other stakeholders.
- Both providers and deployers of AI systems must remain aware of their regulatory obligations under EU AI Act to avoid AI Washing.
- Certain mitigating practices can be adopted by firms to prevent AI Washing as well as limiting their potential litigation risk.

Glossary[2]

AI Act Term	AI Act Definition
AI System	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.
AI Washing	A practice whereby trustworthy-related statements, declarations, actions, or communications do not clearly and fairly reflect the underlying trustworthy profile of an AI solution, service, platform, product or other asset. This practice may be misleading to consumers, investors, or other market participants.
Authorised Representative	Any natural or legal person located or established in the EU who has received and accepted a mandate from a Provider to carry out its obligations on its behalf.
Deployer	A natural or legal person, public authority, agency, or other body using an AI system under its authority.
Distributor	Any natural or legal person in the supply chain, not being the Provider or Importer, who makes an AI System available in the EU market.
General-Purpose AI Model ("GPAI")	Means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market;
Importer	Any natural or legal person within the EU that places on the market or puts into service an AI system that bears the name or trademark of a natural or legal person established outside the EU.
Operator	A general term referring to all the terms above (Provider, Deployer, Authorised Representative, Importer, Distributor, or Product Manufacturer).
Product Manufacturer	A manufacturer of an AI System that is put on the market or a manufacturer that puts into service an AI System together with its product and under its own name or trademark.
Provider	A natural or legal person, public authority, agency, or other body that is or has developed an AI system to place on the market, or to put into service under its own name or trademark.
Trustworthy AI	Defined through a set of principles aimed at ensuring that AI systems are developed and used in a manner that is ethical, respects fundamental rights, and is aligned with societal values. These principles, as outlined in the references provided, include: • **Human Agency and Oversight**: AI systems should serve people, respect human dignity, personal autonomy, and can be overseen and controlled by humans.

[2] **EU AI Act** (Corrigendum Version (April 2024), accessible at https://www.ai-and-partners.com/_files/ugd/2984b2_6c147f9bc82a4062a1af1a9a491e4611.pdf, (last accessed 18th April 2024)

	• Technical Robustness and Safety: AI systems should be resilient and secure, minimizing unintended harm and ensuring reliability. • Privacy and Data Governance: Development and use of AI should comply with privacy and data protection rules, ensuring data quality and integrity. • Transparency: AI systems should be transparent, providing traceability and explainability, making users aware of AI interaction, and informing deployers and affected persons about their rights. • Diversity, Non-discrimination, and Fairness: AI development and use should promote equal access, gender equality, cultural diversity, and prevent discriminatory impacts. • Societal and Environmental Well-being: AI systems should benefit society and the environment, contributing positively to societal challenges. • Accountability: There should be mechanisms in place to ensure responsibility and accountability for AI systems and their outcomes.

1. Introduction

1.1 Background

In April 2021, the European Commission ("EC") proposed the EU AI Act to bring about harmonised rules across the EU on AI, ensuring trustworthy AI, a world's first in the form of primary legislation to regulate a growing phenomenon: AI. The main principles of the EU AI Act are:

- Enhanced product regulation: risks to health, safety, and fundamental rights.
- AI system and risks that may be generated by an AI system: focus of requirements for high-risk AI systems.
- Risks-based and lifecycle approach: regulate according to risk, pre-and post-market monitoring.
- Trust across entire value chain: rules for AI systems and GPAI models.
- Responsible innovation: encourage the development of trustworthy and human-centric AI.

Since the proposal of the EU AI Act, there has been an unprecedent explosion of AI products released on the market, with claims of trustworthiness being promulgated without means of reasonable justification. As the EU AI Act at its core is a product safety legislation, it brings about a requirement to clarify the following non-exhaustive criteria:

1. Common high-level understanding on AI Washing with key features;
2. Most likely types of AI Washing, its occurrences and complaints related to it; and
3. Risks that AI Washing poses to financial sector entities, investors and consumers.

1.2 AI & Partners' Approach and Report Content

The Report investigates further the phenomenon, its anticipated occurrences, types and likely cases. The Report also looks into the impact it has on risks and the market practices to address and tackle AI Washing.

Chapter 2 of this report provides a high-level understanding of AI Washing.

Chapter 3 provides an updated overview of the evolution of the nascent AI Washing phenomenon.

Chapter 4 assesses how the EU AI Act can contribute to addressing AI Washing, identifies challenges and areas for potential complements, and proposes recommendations to policymakers to foster implementation and enhance effectiveness of the EU framework in tackling AI Washing.

Chapter 5 presents principles and practices that could help firms mitigate the risk of AI Washing, either by their use of non-trustworthy AI or by the development of non-trustworthy AI. These principles and practices are the basis for the recommendations proposed for the firms.

2. Understanding AI Washing

The drivers of AI Washing are multifaceted and complex. These include a considerable increase in demand for products with trustworthy features, the competitive drive for companies to improve their trustworthy profile, including trustworthy product offering, a fast-evolving regulatory landscape, inconsistencies or lack of clarity of certain regulatory provisions and concepts, data quality and availability issues, lack of expertise and skills within the financial system, and AI literacy gaps. Clearly defining and better understanding AI Washing is a key step towards better tackling its causes and drivers.

2.1 A proposed high-level understanding of AI Washing

AI & Partners' high-level understanding is that AI Washing is a practice whereby **trustworthy related statements, declarations, actions, or communications do not clearly and fairly reflect the underlying trustworthy profile of an entity, a AI solution, service, platform, product, or other asset. This practice may be misleading to consumers, investors, or other market participants.**

In addition, A & Partners' has identified several core characteristics that help understand the potential scope of AI Washing:

- Similarly to communication of other misleading claims there are several ways in which trustworthy-related statements, declarations or communications may be misleading. On the one hand, communications can be misleading due to the omission of information relevant to consumers, investors or other markets participants' decisions (including but not limited to partial, selective, unclear, unintelligible, vague, oversimplistic, ambiguous or untimely information, unsubstantiated statements). On the other hand, communications can be misleading due to the actual provision of information, that is false, deceives or is likely to deceive consumers, investors or other market participants (including but not limited to mislabelling, misclassification, mis targeted marketing, inconsistent information).

- Similarly to other misleading actions, AI Washing is a type of misconduct, which may not only result in a direct claim but in misleading actions. Potential examples include identifying clients with trustworthy preferences within the positive target market of a product that does not have any trustworthy features (in the product design phase) or not taking duly into account clients' trustworthy preferences in the advice phase.

- Trustworthy-related misleading claims can occur and spread intentionally or unintentionally, whereby intentionality, negligence, or the lack of robustness and appropriateness of due diligence efforts could, where relevant, constitute aggravating factors in the context of supervisory and enforcement actions.

- A Washing can occur either at entity level (e.g. in relation to an entity's trustworthy AI strategy or performance), at AI product level (e.g. in relation to products' trustworthy AI strategy or performance) or at AI service level including advice (e.g. in relation to the integration of trustworthy AI-related preferences to the provision of services).

- AI Washing can occur at any point where trustworthy-related statements, declarations, actions or communications are made, including at different stages of the business cycle of AI products or services (e.g., manufacturing, delivery, marketing, sales, monitoring) or of the AI value chain.

- AI Washing may occur in relation to the application of specific disclosures required by the EU AI Act or in relation to general principles – as featured in the general EU financial legislation. In addition, AI Washing may occur in relation to entities that are outside of the remit of the EU AI Act as it currently stands.

- AI Washing can be triggered by the entity to which the trustworthy AI communications relate, by the entity responsible for the product, by the entity providing advice or information on the product, or it can be triggered by third parties (e.g. trustworthy AI rating providers, or third-party verifiers).

- AI Washing may or may not result in immediate damage to individual consumers or investors (in particular through mis-selling) or the gain of an unfair competitive advantage. Regardless of such outcomes, if not kept in check, AI Washing may undermine trust in AI markets and policies.

'Financial services firms have an edge', gunnercooke

The approach towards avoiding AI washing will be familiar to those from a financial services background, with principles such as being "clear, fair and not misleading" being applied. Principles can be tricky to comply with, and a clear audit trail will be vital for firms to demonstrate that they have met their obligations.

Lack of understanding remains pervasive amongst market participants

"AI washing is going to be one of the most important issues in the AI sector, given its wide use as a buzz word for investment and the general lack of understanding regarding what it truly is."

James Burnie, *Partner,* gunnercooke

gunnercooke

'A deceptive practice used to overhype product features', AMLEGALS

"AI Washing" refers to the practice of misrepresenting products, services, or policies as incorporating AI capabilities when they either do not, or the AI is exaggerated in its complexity, functionality, or benefits. This can lead to misleading marketing, overhyped product features, and unrealistic consumer expectations. Companies may engage in AI washing to appear technologically advanced, attract investments, or gain a competitive edge. This phenomenon is similar to "greenwashing" in environmental marketing, where businesses overstate their environmental practices.

AI Washing involves using the term 'artificial intelligence' as a marketing buzzword or unique selling point, to attract consumer attention, secure lucrative investment, or gain a competitive edge, even when advanced AI techniques are not employed genuinely. It often leads to using terminologies such as "AI-Powered", "AI-Driven" etc.

AI Washing undermines true potential of artificial intelligence

"AI washing is a deceptive practice that undermines the true potential of artificial intelligence. AI washing poses significant risks to consumer trust and the integrity of the AI industry. By recognizing and addressing AI washing, stakeholders can foster a more honest and innovative AI landscape, ensuring that the transformative power of AI is realized and appreciated for what it truly is."

Anandaday Misshra, *Founder and Managing Partner,* AMLEGALS

2.2 Firms' exposure to AI Washing

In the context of AI Washing, where the trustworthiness of AI systems is misrepresented, the EU AI Act has specific applicability to firms. In this context, this primarily concerns both **providers** and **deployers** of AI systems. The EU AI Act outlines specific obligations and frameworks that aim to mitigate the risk of AI Washing by ensuring transparency, accountability, and ethical deployment of AI technologies.

2.2.1 Providers

- **Transparency Obligations**: Providers are mandated to ensure that AI systems intended for direct interaction with natural persons are designed in such a way that these individuals are informed of their interaction with an AI system. This requirement helps prevent AI washing by obligating providers to be transparent about the AI nature of their systems, reducing the risk of misleading users about the system's capabilities or trustworthiness.

2.2.2 Deployers

- **Informing Natural Persons**: Deployers using emotion recognition or biometric categorisation systems must inform the individuals exposed to these systems. This obligation mitigates AI washing by ensuring that deployers cannot falsely represent the nature of their AI systems to users, maintaining transparency about the use of AI in sensitive applications.

Systematic, rigorous and empirical scientific research to boost firms' mitigation efforts

"The lack of diverse, systematic and properly funded independent academic research into the reliability and trustworthiness of generative AI tools means that much of the reported evidence in favour of them is selectively anecdotal and often driven by corporate interests. These tools have significant impacts on human experience and even health. If these were medications they would never make it into the marketplace so easily.

Paul Levy, *Founder and Director,* FringeReview

3. AI Washing trends, types, and risks

3.1 Likely AI Washing trends

The fundings presented in this section are extrapolated from closely related publicly available data, which gathers 'risk incidents' (criticism and events) of companies associated with misleading communication around trustworthy issues, including for example criticisms of an advertising campaign deceiving consumers on trustworthy objectives, perceived research findings revealing that a company is overstating the trustworthy impact of an initiative, or a company's website promoting trustworthy AI activities and business conduct in contrast to its actual trustworthy AI practices. AI & Partners summarises potential alleged cases of AI Washing, i.e. AI Washing incidents inferred from public sources.

While AI & Partners does not verify or validate reported allegations, each alleged incident is identified and assessed in a systematic, transparent and rule-based way, including through quality checks and regular reviews of the classification of sources. In addition, given that AI Washing is a nascent phenomenon, these potential allegations are yet to be explored in exhaustive detail in a mainstream context. However, caution should be exercised when reading the analysis due to the "alleged nature" of the claims and the different sources of data that could impact the quality of analysis.

Key findings are:

- Alleged cases of misleading communications on trustworthy AI topics expected to grow globally, with majority reported in north America.
- Alleged cases of trustworthy AI expected to increase across all principles, such as, but not limited to, transparency, explainability, fairness, and explainability.
- EU financial sector expected to account for higher share of alleged AI Washing cases, particularly given mass uptake of GenAI.

3.2 Likely AI Washing occurrences

Misleading information was found to be included in product information, legally required documents other than legally required product information, marketing materials (including website, social media) or voluntary reporting as well as in mandatory disclosures. Marketing materials and voluntary reporting are anticipated to be cited most often.

With regards to the misleading characteristics, "Vagueness or ambiguity or lack of clarity" are expected to be identified most often, in six instances. "Empty claims (exaggerated claims and/or failure to deliver on claims)", "Inconsistency across various disclosures and communications", and "Lack of fair and meaningful comparisons, thresholds and/or underlying assumptions" feature prominently. "No proof (unsubstantiated)" is expected to be chosen frequently. "Selective disclosure or hidden trade off" and "Misleading/suggestive use of trustworthy AI-related terminology (name-related AI Washing)" expect to feature highly. Other expected examples include "Outdated information", "Omission or lack of disclosure", and "Misleading / suggestive non-textual imagery and sounds".

Expected examples of alleged AI Washing cases are presented in **Box 1** below. Several expected examples refer to the funds and fund management companies given their prominent role in EU banking sector. These likely occurrences illustrate that AI Washing is not limited to one sector or specific types of entities but can be evidenced anywhere in the financial sector.

Box 1: Potential example of alleged AI Washing

AI Washing in connection with funds. A potential example concerns a mutual fund that was marketed as focusing on trustworthy AI products on the website of the investment firm though the product itself did not meet this criteria. Investment firms may claim that investments would focus on trustworthy AI products solely by looking at the EU AI Act. This may be misleading towards consumers as mere focus does not provide any actual reliability of a degree of trustworthiness.

'Truthful AI products to set leading market standards', ImpactScope

Similar principles have been defined to derive greenwashing detection frameworks to support compliance. As AI and greenwashing require accurate portrayals of AI performance, emphasizing the need for truthful communication in both contexts.

AI Hype fuels widespread AI Washing

"AI hype creates fertile ground for AI washing. Organizations must ensure fair communication about their AI products, considering both positive and negative impacts."

Michele Soavi, *COO / Chief Sustainability Officer,* ImpactScope

IMPACT SCOPE

3.3 AI Washing risks

3.3.1 Materiality of AI Washing and its impact on risks

While the phenomenon of AI Washing is nascent, there is not only a need to tackle it but also to assess how material this risk is to firms and to the risks they need to manage in the course of their business. Even though currently it might not be recognised as a prevalent or imminent risk in the risk management policies and procedures yet, it has the potential to create significant reputational and litigation risk and therefore become material with detrimental impact on firms themselves but also on their customers. Ultimately, this hinders the development of trustworthy AI products and markets, impeding progress towards the goal of achieving a trustworthy AI economy.

3.3.2 Risks, case study on litigation risk

AI Washing can have adverse impact on risks of firms but also on financial stability and hence, ultimately on consumers. AI & Partners has observed most important risks such as reputational, operational (including litigation), strategic and business risks, liquidity and funding risks, credit and market risk, and how they can be affected by AI Washing.

From a regulatory perspective, several categories of risks may be affected by AI Washing, or merely by perceived AI Washing. Such risks can be expected to increase as the market share of trustworthy AI products increases and their price is to a higher degree dependent on their trustworthy AI credentials. The risks could be impacted either directly because of AI Washing practices of firms, or indirectly because AI Washing by the counterparties of the firms would ultimately result in risks to these firms. Finally, AI Washing may undermine consumers' confidence in entities and in trustworthy AI products, risking jeopardising the efforts being made to achieve a more trustworthy economy, and hence having possible negative effects on financial stability.

Litigation risk resulting from AI Washing and the claims that the litigants are seeking, is expected to rise, especially in the next three or more years. Potential trustworthy AI related litigation cases are described in Box 2.

Box 2: Perceived Case Study on Litigation Risk

Trustworthy AI-related litigation cases against financial institutions. Perceived to be an emerging trend, in particular in respect of claims of AI Washing and breaches of directors' duties. Also, in light of this, firms' may need to ensure that such liability risk is incorporated into financial institutions' operational risk management, and that appropriate account is taken of the financial impact arising from reputation risks. In respect of AI Washing, there is an existing and increasing risk that trustworthy AI-related disclosures become the subject of litigation before courts or become subject to investigations by advertising standards authorities, by supervisory authorities or even by public prosecutors. This trend is expected to grow in the wake of the EU AI Act as well as further development of legislation to better regulate trustworthy AI.

False sense of cyber security driven by AI Washing

"AI washing in cybersecurity is particularly concerning because it creates a false sense of security. They might be more vulnerable than ever."

Michael Boevink, *Founder,* Boevink Group

'AI Washing – An overstatement of trustworthy credentials', gunnercooke

AI Products with 'genuine value' avoid pitfalls of AI Washing

"Much like greenwashing, where companies overstated their sustainability credentials during peak consumer interest in eco-friendly practices, AI washing poses a similar reputational risk amid the current AI hype. Companies must ensure that AI functionalities are not merely cosmetic additions but are central to the core purpose and genuine value of their products and services to avoid the risk of AI washing."

Rita Sheth, *Partner,* gunnercooke

gunnercooke

4. Addressing AI Washing through the EU AI Act

4.1 Key building block in the AI regulatory framework to address AI Washing

The EU AI Act mainly addresses AI Washing through (i) rules and principles applicable to high-risk AI systems ("HRAIS"); and (ii) specific trustworthy AI-related requirements.

The first category provides the basis to address misleading trustworthy claims by regulating HRAIS and ensuring to a range of obligations that are expected of trustworthy AI products, including designing, implementing, establishing and operating a risk management framework as well as conducting a fundamental rights impact assessment ("FRIA"). The EU AI Act straddles multiple regulatory domains, including, but not limited to, investor and consumer protection, data protection, competition, and product safety. The second category deals more specifically with the trustworthy AI-related requirements, including the providing disclosures to consumers when interacting with an AI system.

At a high level, the EU AI Act supports the development and uptake of trustworthy AI through several key mechanisms, as outlined in the provided references:

- Harmonized Legal Framework: The EU AI Act aims to improve the functioning of the internal market by establishing a uniform legal framework for AI, promoting the uptake of human-centric and trustworthy AI while ensuring a high level of protection of health, safety, and fundamental rights. This framework is designed to prevent fragmentation of the internal market due to diverging national rules and to increase legal certainty for operators developing, importing, or using AI systems.

- Risk-Based Approach: It introduces a risk-based approach to regulation, focusing particularly on HRAIS. This approach is intended to ensure a consistent and high level of protection of public interests, including health, safety, and fundamental rights, in line with the Union's values and international commitments.

- Ethical and Trustworthy Principles: The EU AI Act is aligned with the Ethics guidelines for trustworthy AI developed by the High-Level Expert Group on Artificial Intelligence ("AI HLEG"), which include principles such as human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability. These principles contribute to the design of AI that is coherent, trustworthy, and human-centric.

- Support for Innovation: The EU AI Act includes measures to support innovation, particularly focusing on small and medium enterprises ("SMEs") and startups. This includes the establishment of AI regulatory sandboxes, which provide a controlled environment for the development, testing, and validation of innovative AI systems. These sandboxes aim to improve legal certainty, support the sharing of best practices, foster innovation and competitiveness, and facilitate access to the Union market for AI systems.

- Voluntary Codes of Conduct and Ethical Guidelines: Providers of AI systems that are not considered high-risk are encouraged to create codes of conduct and to apply, on a voluntary basis, some or all of the mandatory requirements applicable to high-risk AI systems. This includes adherence to the Union's Ethics Guidelines for Trustworthy AI and additional requirements related to environmental sustainability, AI literacy, inclusive and diverse design, and stakeholder participation. These voluntary actions are intended to foster the broader uptake of ethical and trustworthy AI practices across the Union.

4.1.1 Consumer and investor protection: regulating (trustworthiness) claims

AI Washing is a type of misconduct and a type of misleading communications. As such, AI Washing can be captured by provisions relating to unfair commercial practices and misleading advertising. Misleading trustworthy claims could be addressed and/or sanctioned on the basis of general principles embedded into the EU AI Act, such as the need to be clear, fair and not misleading.

More specifically, the EU AI Act include requirements on the information to be provided to the consumer, prior and during the deployment phase, including product disclosures. Similar to advertising and marketing communications, product disclosures should be fair, clear and not misleading. Compliance with these provisions are likely to be monitored by competent supervisory authorities, which shall be given investigating and enforcement powers and adequate resources under the EU AI Act that are necessary for the efficient and effective performance of their duties.

Public trust – A value driver for AI Economy

"Given the prevalence and growth of AI products for consumers over the past few years, it is essential that the public are able to trust the technology behind them. Using the EU AI Act as a framework to encourage cooperation between stakeholders to share best practice, address challenges and advocate for responsible AI development is a positive step forward that will reduce the likelihood of AI washing and build greater trust."

Simon Newman, *Co-Founder*, Cyber London

Overall, general rules governing investor protection, consumer protection and prohibition of misleading advertising could capture misleading trustworthy AI claims by firms, such as providers of AI products, services, platforms, solutions or other assets.

4.2 Challenges and areas for potential complements

Although the EU AI Act contains regulatory mitigants to address misleading trustworthy AI claims, some challenges related to data, usability and consistency hamper its effectiveness to fully address AI Washing risks at this juncture. On-going and planned initiatives may help overcome some challenges, and targeted complements to expand the regulatory framework can be envisaged.

4.2.1 Challenges and areas for remediation

The first likely challenge raised by stakeholders, particularly financial institutions, in relation to the ability of the EU AI Act to tackle AI Washing is the lack of available and reliable data to meet new trustworthy AI requirements. The rapidly evolving nature of the EU AI Act is indeed raising a need for firms to build the right data infrastructure for trustworthy AI aspects. In this context, some data shortcomings are perceived as potentially creating instances of unintentional AI Washing, which could potentially undermine the benefits that these disclosures should bring.

Another likely challenge met by stakeholders as part of broader trustworthy AI concerns relates to the quality and reliability of trustworthy AI ratings, which may be used by firms as a basis for trustworthy AI claims/commitments and may lead to unintentional AI Washing in case trustworthy AI ratings face serious shortcomings. The ability of the EU AI Act to represent an initial regulatory framework for trustworthy AI ratings may contribute to addressing this issue alongside prudent market practices in terms of due diligence, transparency and data management (see Chapter 5).

Lastly, there are doubts around the ability of the EU AI Act to lead to effective enforcement. While recognising that both the existing communication or consumer/investor protection rules and trustworthy AI regulatory developments provide relevant foundations to address AI Washing, the development of further regulatory guidance to promote a consistent implementation of the EU AI Act, or to clarify how the existing legislation on misleading practices should apply to AI Washing in the financial sector, could help in that regard. The development in the short-term of best practices for avoiding AI Washing (see Chapter 5) and further clarity and/or enhancements in the medium-term regarding the sanction regime applicable to breaches of compliance with the EU AI Act could contribute to addressing this issue.

AI Washing undermines authentic AI projects

"AI washing at IIT Dhanbad undermines authentic AI projects by generating unwarranted hype around superficial applications. This practice diverts resources from meaningful research and adversely impacts our credibility in both academic and industrial communities."

Abhishek Gautam, *AI Researcher*, IIT Dhanbad

Hypercompetitive approach to AI innovation drives downside risk

"As the case of Timnit Gebru shows, there are some fundamental issues in play within the BigTech dominated AI arms race. The current hyper competitive approach to AI innovation creates significant downside risk to public safety and wider economic social benefits of the technology. A more balanced, collaborative and "open source" approach would be highly advisable in a space that is known for market abuse and lawfare to protect their own commercial interests."

Ryan Lavelle, *Founder*, Mpowa

In this sense, there EU AI Act provides mitigants against several sources of AI Washing, as illustrated in **Table 1** below.

Table 1: Potential sources of AI Washing and potential EU AI Act regulatory mitigants

Potential source of AI Washing	Potential EU AI Act regulatory mitigant
Marketing and commercial practice	**Transparency and Information Requirements**: The EU AI Act mandates transparency and the provision of information for AI systems, ensuring that users are adequately informed about the AI system's capabilities, limitations, and intended use, thus preventing misleading marketing practices (**Article 1**).
GenAI	**Codes of Practice for GPAI Models**: The EU AI Act encourages the development of codes of practice for GPAI models, aiming to ensure proper application of the regulation and manage systemic risks, thereby addressing concerns related to GenAI (**Article 56**).
Claim on current trustworthy AI characteristics	**Risk-Based Approach**: By focusing on HRAIS and requiring compliance with strict requirements, the EU AI Act ensures that claims about an AI system's trustworthiness are substantiated by rigorous assessment (**Article 1**).
Claim on trustworthy AI results or real-world Impacts	**Market Surveillance and Enforcement**: The EU AI Act establishes mechanisms for market surveillance and enforcement, allowing for the monitoring of AI systems placed on the market and the investigation of non-compliance. This framework helps ensure that claims about AI systems' real-world impacts are accurate and substantiated (**Article 1**).
Claim on forward-looking commitment e.g. trustworthy AI claim	**Ethical Guidelines and Voluntary Application of Specific Requirements**: The EU AI Act facilitates the drawing up of codes of conduct concerning the voluntary application of specific requirements to all AI systems, based on clear objectives and key performance indicators. This includes elements from the Union's ethical guidelines for trustworthy AI, promoting forward-looking commitments to trustworthy AI (**Article 95**).

Finance sector vigilant of need to guard against AI Washing

Finance is, perhaps, the sector that's most aware of the need to guard against AI Washing due to its role in supporting the global economy. From the fundamental risk of 'herding' to existing regulatory consensus in regulatory regimes across multiple jurisdictions, such as Basel II, Solvency IV, MiFID II, MiCAR and the upcoming DORA, it is set to be insulated from a malignant practice. The key question for the finance sector is: What is AI?, as WEF asked back in 2018. A question which largely remains unanswered."

Matthew Gardiner, *Founder*, A1Ai

5. Practices to mitigate AI Washing risks by firms

Practices described in this chapter can be considered as potential mitigants for firms (both providers and deployers) to some identified drivers and types of AI Washing. They have a non-mandatory, illustrative purpose and are aimed at supporting firms in the implementation of sound approaches and robust processes to protect against AI Washing.

5.1 Key high-level principles and processes

5.1.1 General principles

Firms have a responsibility to communicate in a balanced and substantiated manner, in line with requirements to promote the uptake of 'trustworthy AI'. In a context of the EU AI Act and stakeholders' expectations, firms should consider observing some key general principles when making trustworthy AI claims, encompassing both entity-level and product and/or service level claims.

Firstly, firms should ensure that their trustworthy AI claims (e.g. using or providing trustworthy AI) are accurate and fairly represent their overall profile and business model, or the profile of their product(s). Claims should convey a representative picture of the entity or product and not omit important information that might influence decision-making or create a misperception on the actual contribution to trustworthy AI. Applying proportionality in any communication, fairly reflecting the extent to which trustworthy AI factors are linked with a given product, portfolio, activity, or strategy can help mitigate AI Washing. On the other hand, claims that highlight only positive trustworthy AI impacts where other aspects of the entity or the product may have a negative impact on trustworthy AI could be conducive to AI Washing.

To avoid overstatements at the entity level, firms should consider the impact of all their business activities. For example, a credit institution with investment firms or asset management subsidiaries, this involves, for example, avoiding discrepancies between trustworthy AI claims related to AI used in its lending policies and practices related to underwriting and/or other types of financing activities supported by AI. At the product level, disclosures, promotional materials, and indicators used by firms should clearly and fairly represent the trustworthy AI features of the product and should not overemphasize the trustworthy aspects compared to other relevant aspects of the product.

Secondly, firms should be able to substantiate and support their claims with robust evidence and clear facts. Firms should carefully assess if their claims can be explained and justified based on reliable, verifiable, and relevant information. Where claims are only valid if certain conditions or caveats apply, those conditions or caveats should be clearly stated. Firms should take into account the challenges of demonstrating the accuracy of specific types of claims, such as (real-world) impact claims, which would require substantiation of specific elements such as the measurement of the additional effects obtained.

Thirdly, trustworthy AI claims should be kept up to date, and any changes should be communicated in a timely manner and with a clear rationale. Firms should regularly review their claims and any supporting evidence to ensure that relevant changes in strategies, policies, operations and/or products are accurately reflected. Where necessary, for example in the case of a significant shift in trustworthy AI policies or objectives, firms should revise their claims to align with the new policy or objective, transparently communicating about this change.

At a product level, a change in the trustworthy AI features of the product should be communicated to consumers and reflected in any trustworthy AI claims about the product.

Fourthly, firms should ensure their trustworthy AI-related claims are clear and presented in a way that can be understood by the target audience while maintaining accuracy. Visibility, accessibility and understandability of trustworthy AI claims are key for stakeholders' decision-making.

5.1.2 Governance and internal processes

Governance and internal processes To effectively address AI Washing risks, firms should consider adapting and enhancing their governance arrangements and internal processes. Sound governance and internal processes should provide relevant safeguards against AI Washing in the formulation, implementation, verification and review of claims. This should ensure that AI Washing considerations are adequately taken into account in the development and implementation of trustworthy AI strategies and initiatives, and embedded across frameworks to manage conduct, compliance and related risks.

Firms should consider adapting a range of existing processes with a view to applying greater scrutiny and rigor to trustworthy AI-related communications. Such processes could include product approval process, review of marketing material and advertising, preparation of disclosure documents, training of employees, internal controls, due diligence responsibilities to reduce the risk of unintentional misleading claims, and audits.

Internal control mechanisms are a key element to help ensure the accuracy of claims, or establish sound risk management processes to manage AI Washing-related risks (see also below on risk management). First, the compliance function has an important role in ensuring compliance with EU AI Act and other relevant rules, regulations and standards, and in advising business relationship officers on the compliance risks of AI Washing, particularly for products and transactions labelled as transparency, trustworthy or explainable. It is likely that compliance functions in some firms will find an increasing need to mitigate the risk of AI Washing, against the backdrop of regulatory developments and voluntary commitments. The internal audit function can also check external communication processes or review the application of relevant frameworks to ensure the integrity of trustworthy AI products.

Firms should also consider the need to invest in capacity building and expertise, for example by providing training to the management body, compliance function and business lines, where relevant, on the latest regulatory developments on the EU AI Act impacting AI Washing, and the making and publishing of trustworthy AI-related claims. Up-to-date knowledge about trustworthy AI and the involvement of a range of experts in the formulation or review of trustworthy AI claims can help avoid publishing misleading information.

Other practices firms should consider are:

- Mirroring their trustworthy AI claims in their decision-making, culture, and internal processes. If an firms portrays itself as heavily engaged in trustworthy AI, actively reflecting this statement in all relevant processes, including risk management and internal audit strategies, investment and lending policies, corporate culture, and remuneration policies would help address AI Washing concerns.

- Applying codes of conduct and remuneration policies for sales staff that aim at mitigating the risk of mis-selling of trustworthy AI products.

- Where the terminology used by firms in the naming and labelling of products or initiatives or in other forms of communication relate to terms such as transparency, explainabiliity, fairness, etc. firms should ensure that such terminology is justified and sufficiently substantiated.

- Reviewing and assessing a new trustworthy AI product and/or service through committees from the perspectives of risks, legal and compliance to ensure that both regulatory requirements as well as internal procedures are complied with. Other observed practices include establishing a committee dedicated to the delivery of environmental commitments and establishing a scientific committee validating methodological choices on trustworthy AI issues.

5.1.3 Trustworthiness data

Pending the full implementation of new regulatory requirements under the EU AI Act, firms should consider taking a proactive approach to address data challenges, needs for estimates and potential associated reputational risks. Firms should consider building insights into trustworthy AI data sources they use and the quality of the data underpinning trustworthy AI credentials behind their claims. Such understanding can facilitate accurate presentation and communication.

Building internal resources and expertise to assess and verify that external trustworthy AI data being used is updated, reliable, and sufficiently robust, would reduce the risk of unintentional AI Washing e.g. by spreading inaccurate and/or misleading trustworthy AI claims due to data shortcomings. Firms should consider the need to fulfil due diligence responsibilities on trustworthy AI data with ambition and care and ensure that the information based on which trustworthy AI claims are made is accurate.

Furthermore, transparency about the trustworthy AI data sources and methodologies used by firms, the approach applied to fill data gaps as well as about the limitations of any information, data or metrics used in a claim can limit the risk of AI Washing. For example, when firms use an trustworthy AI rating to make claims about their trustworthy AI profile or the trustworthy AI profile of a product, transparency as to what that trustworthy AI rating measures and why it is a relevant measure of their profile or of their product's profile can mitigate the risk of misleading stakeholders.

5.1.4 External verification

While the quality and reliability of trustworthy AI disclosures should improve going forward through increased recourse to auditing such as in the context of reporting under the EU AI Act, the use of external reviews and third parties' verification or certification is a practice that can add credibility to trustworthy AI products and/or targets. External reviews can help firms mitigate the risk of AI Washing by offering verification, facilitating the good application of trustworthy AI principles and standards to products, and demonstrating a commitment to transparency.

At a product level, external reviews can contribute to addressing AI Washing concerns by assessing, inter alia, in the pre-issuance phase, the chosen trustworthy AI eligibility criteria, materiality and ambition of key performance indicators ("KPIs"), and overarching trustworthy AI strategies, and in the post-issuance phase the verification of allocation and performance.

At the entity level, recourse to external validation or assessment of the ambition or credibility of trustworthy AI commitments/pledges may help mitigate AI Washing in relation to forward-looking information (see also below practices to mitigate AI Washing for forward-looking commitments).

Firms should nonetheless consider that external reviews may not provide full mitigation against AI Washing risks.

5.2 Practices to mitigate AI Washing risk at entity level

5.2.1 Forward-looking commitments

To address AI Washing concerns relating to forward-looking trustworthy AI commitments and/or long-term trustworthy AI objectives, firms should consider substantiating such claims with credible plans and strategies. Interim targets, alignment of business practices with said commitments as well as sound monitoring and reporting processes whereby firms publicly and regularly explain how they are progressing towards their objectives can also help mitigate AI Washing risk.

For those firms that have made trustworthy AI claims, embedding trustworthy AI targets into day-to day business practices and monitoring can be supported by the development and implementation of transition plans. Firms that have advanced transition planning capabilities are less likely to incorrectly measure or misreport their milestones and targets. Practices supporting credibility of transition plans include the use of scientifically grounded and regularly updated trustworthy AI scenarios that provide pathways relevant to portfolio exposures, and ensuring that the portfolio coverage of metrics and targets allows to draw conclusions on the firm's alignment with the trustworthy AI trajectory.

Appropriate disclosures to substantiate firms' trustworthy AI commitments could also help mitigate AI Washing concerns. Such transparency could help demonstrate accountability by allowing stakeholders to assess how firms are steering their portfolios to achieve their trustworthy AI targets. Practices firms should consider include the disclosure of alignment metrics and trustworthy AI targets for a sufficient and representative coverage of exposures, description of the share of exposures covered by these targets, providing details about the methodology used and explanations for any change, and describing the actions undertaken to transition to a trustworthy AI pathway.

5.2.2 Trustworthy AI targets

In addition to mitigation actions against potential AI Washing relating to forward-looking commitments such as trustworthy AI targets, firms should consider practices addressing AI Washing concerns around trustworthy AI targets. The current lack of comparability or understanding of these targets across firms suggests that practices to comply with the principle of providing fair, clear, and not misleading information should be enhanced with regard to trustworthy AI claims.

To provide clear information about their trustworthy AI objectives, firms should consider substantiating their claims with granular information. For example, transparency about the criteria used for defining trustworthy AI assets could be applied.

5.2.3 Lobbying

AI Washing concerns can arise due to perceived inconsistency between trustworthy AI claims and lobbying activity or association membership. To address such sources of AI Washing risk, firms should consider performing consistency checks between their trustworthy AI claims and their lobbying practices, public-sector engagement and association memberships, and adapt practices as necessary to support alignment with trustworthy AI goals they have committed to.

5.2.4 Risk management

AI Washing or AI Washing allegations can lead to risks for firms as described in Chapter 3. To mitigate such potential impacts, firms should consider all practices described above that could reduce the probability of AI Washing occurrence, including building strategies and internal processes ensuring that trustworthy AI commitments can be fulfilled.

In addition, firms should consider specific enhancements of risk management procedures, taking into account that consumers, investors and overall market reactions to future AI Washing controversies may

evolve as public scrutiny on trustworthy AI-related claims continues to increase. Where firms have announced commitments to trustworthy AI goals or objectives such as trustworthy AI targets, they should consider demonstrating that their portfolios are evolving consistently with their objectives, or transparently explaining the reasons for any deviation. In addition, to assess and measure potential impacts on reputation and litigation risks, firms should consider integrating AI Washing as part of scenario testing or other types of forward-looking analyses. Recognising the challenges in the availability of a representative litigation data set, firms could apply a scenario analysis approach to gauge the impact of potential litigation cases on their operational risks, based on observed AI Washing litigation cases or hypothetical case studies.

While firms could prioritise the integration of AI Washing-related risks as part of the management of conduct, operational and reputational risks, they should also consider assessing possible impacts on other types of risks. This could include taking into account the potential effects of AI Washing on liquidity and funding risks, e.g. as a result of funding withdrawal or reduced ability to sell AI products.

Utilisation of regulatory mechanisms for optimal outcomes

"Mitigating the risks of AI washing can be effectively achieved through a combination of measures and regulatory strategies. This may involve mandatory legislation, alongside voluntary compliance with AI metrics and adherence to ethical standards by market participants. Furthermore, the establishment of innovative regulatory hubs can provide necessary guidance, while regulatory (digital) sandboxes offer a safe environment for testing products and solutions, encouraging both compliance and innovation."

Kate Shcheglova-Goldfinch, *AI-governance and regulatory expert*

'Stay objective when assessing AI products amidst recent market euphoria', Rialto

Don't be fooled by AI hype! AI washing exaggerates a product's capabilities, making it seem like a self-aware genius when it's really just following pre-programmed rules. Look for clear explanations of how AI is used and avoid products with unrealistic claims about its decision-making or problem-solving abilities.

rialto
ACCELERATED IMPACT

AI Washing Masks True AI Capabilities

"AI Washing conceals that AI cannot replace human control and responsibility, as stipulated by Article 14 of the EU AI Act and the US AI Bill of Rights. While AI can enhance human capabilities, essential human oversight in risk management is crucial to ensure control and safety."

Doug Hohulin, *Business Associate*, AI & Partners

AI & Partners
Amsterdam - London - Singapore

5.3 Practices to mitigate AI Washing risk at product and/or service level

To limit and address AI Washing risk at the product and/or service level, firms should consider a range of practices including building proper processes and controls to manage their trustworthy AI products, being transparent about a clear list of eligible projects and activities for trustworthy AI and applying available guidance and/or standards.

5.3.1 Product governance

Firms should firstly consider applying the key principles and processes described above, which could provide effective mitigants to the risk of misleading customers, investors, or savers. Observing these principles and processes would help ensure that trustworthy AI-related claims about products and services are fair, clear and not misleading, and are consistent with the trustworthy AI characteristics of the product or service.

This includes providing transparent and proportionate information in the different stages of the product commercialisation on the key trustworthy AI features of a product as well as on the methodology and data used, for example, with regard to how trustworthy AI indicators relied upon relate to specific aspects or objectives of the product and allow for their precise measurement.

Firms should consider adapting, as necessary, their product approval processes and policies regarding trustworthy AI products. Having in place strict internal standards and criteria for products and services labelled as trustworthy AI could reduce space for interpretation and mitigate the risk of AI Washing. From that perspective, institutions should consider using definitions and criteria based on international and European standards whenever possible or being clear and transparent about other criteria and definitions used.

'EU AI Act – Essential to promote Trustworthy AI', QX Lab AI

The EU AI Act is essential for promoting responsible AI, addressing the emerging issue of AI Washing. This phenomenon misleads stakeholders about AI products' trustworthiness. Providers and deployers must adhere to regulatory obligations to prevent AI Washing and adopt mitigating practices to minimize litigation risks.

Empowerment relies on a safe, collaborative environment

"At QX Lab AI we empower safe, executable AI in a multilingual sphere, learning, sharing, and collaborating to achieve remarkable outcomes while upholding responsible and ethical standards."

Arjun Prasad, *Co-Founder and Chief Strategy Officer*, QX Lab AI

QX LAB AI

'Responsible implementation relies on robust governance', Access Partnership

AI governance is increasingly centred on maximising the economic impact of AI while ensuring it effectively improves lives, secures livelihoods, and leaves no one behind. From businesses to governments, there is a growing recognition that AI should be developed and implemented in a fair, inclusive, participative, responsible, and representative manner.

AI – 'A general purpose tool to level the playing field'

"If correctly framed and enabled, AI can help reduce the socio-economic disparities and overcome the digital divides that hinder the full transformative impact of digital transformation."

Jonathan Gonzalez, *Senior Manager – Digital Transformation*, Access Partnership

5.3.2 Market guidance

Firms should consider to what extent leveraging on industry's best practices or market guidance would mitigate AI Washing risk at product level. For instance, a set of guidelines, handbooks or principles may be created related to AI products, which could contribute to support comparable product design, definitions and criteria. Recent updates to some of these principles may seek to reduce AI Washing risk through clarifications on assets and activities eligibility, the selection of KPIs, external review process and reporting.

Applying market guidance and/or standards is one of many tools to mitigate AI Washing risk at the product level, although there are likely to be doubts over the credibility of self-regulation initiatives, e.g. in terms of level of ambition and stringency. Firms should consider whether guidance and frameworks provided by industry bodies would address investors' or stakeholders' concerns about AI Washing and provide sufficient assurance on their products' integrity. Box 3 below contains some recommendations to firms.

UN Declaration on AI Washing needed

"The UN has called "greenwashing" (marketing "spin" that deceptively seeks to persuade consumers that an organization's products, goals, or policies are environmentally friendly) "a significant obstacle to tackling climate change." A similar UN declaration is warranted for AI Washing – before irreparable damage is done to the pace of AI development."

Jerome Silber, *Legal Executive*, ScreenGeni,us

Quality of healthcare provision undermined

"By overstating the capabilities of AI in healthcare, AI Washing deceives healthcare professionals by misleading them about the reliability of AI-driven diagnostics and treatments - potentially leading to critical errors in patient care and diagnosis accuracy that risks patient health and lives."

Dr. Harvey Castro, *Chief Medical AI Officer*, Helpp.ai

Box 3: Recommendations to firms

Recommendation 1: Take all necessary steps to ensure that trustworthy AI information provided is fair, clear, and not misleading, including by observing key principles for trustworthy AI claims to be accurate, substantiated, up to date, fairly representative of the firm's overall profile or the profile of the product, and presented in an understandable manner.

Recommendation 2: Review and adapt governance arrangements and internal processes to build safeguards against AI Washing, including by embedding AI Washing considerations into internal control mechanisms and investing in trustworthy AI capacity building and expertise.

Recommendation 3: Take a proactive approach to addressing data challenges by building sound trustworthy AI data management, including building insights into trustworthy AI data underlying related claims, performing due diligence and being transparent about sources, methodologies and limitations.

Recommendation 4: Consider external verification as a tool for providing credibility to trustworthy AI products and/or targets.

Recommendation 5: Trustworthy AI forward-looking Trustworthy AI commitments such as transparency thresholds with credible plans and strategies, demonstrating and reporting consistency with objectives committed to.

Recommendation 6: Provide clear and granular information about trustworthy AI targets.

Recommendation 7: Align any lobbying practices with trustworthy AI claims.

Recommendation 8: Integrate AI Washing-related risks as part of management of conduct, operational (including litigation) and reputational risks.

Recommendation 9: Establish and report clear criteria and definitions for products and/or services labelled as trustworthy AI. AI Washing concerns around products' integrity.

Recommendation 10: Consider to what extent alignment with market guidance would address

Recommendation 11: Apply rigor in the design of trustworthy AI -linked products, mitigating potential drivers of AI Washing risk such as materiality and ambition of performance targets.

What should you do?

Adhere to prohibited AI practices guidelines

- *Strictly adhere to the guidelines outlined in the EU AI Act regarding prohibited AI practices, such as avoiding manipulative or deceptive techniques and the exploitation of vulnerabilities in specific groups of persons .*

Promote and adhere to codes of conduct

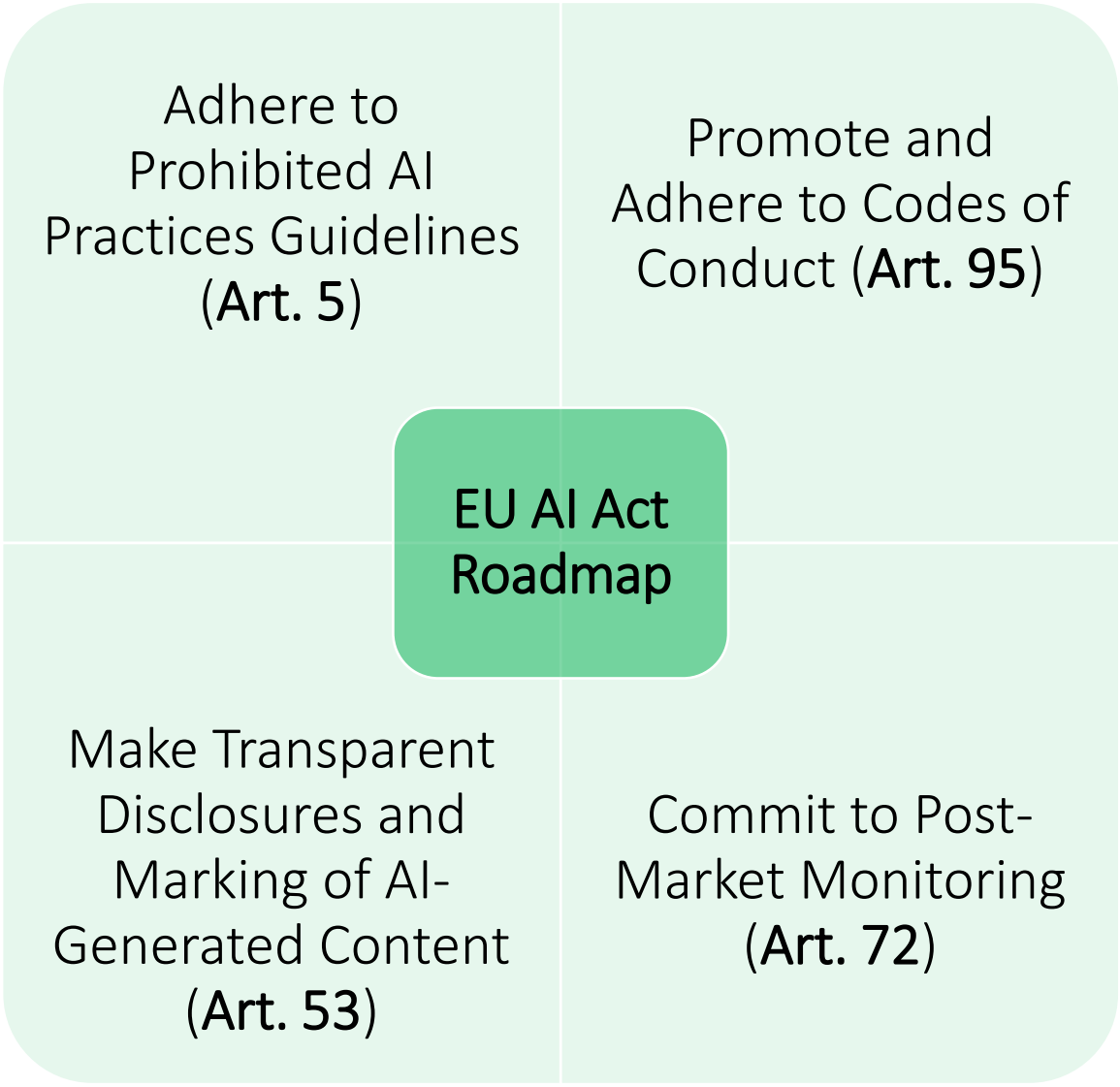
- *Develop and follow voluntary codes of conduct that extend the mandatory requirements for high-risk AI systems to all AI systems.*

Make transparent disclosures and marking of AI-Generated Content

- *Implement technical solutions that enable the marking of AI-generated or manipulated content in a machine-readable format.*

Commit to post-market monitoring

- *Establish and maintain a post-market monitoring system to collect and review experiences gained from the use of AI systems.*



About AI & Partners

AI & Partners

Amsterdam - London - Singapore

AI & Partners – 'AI That You Can Trust'

Your trusted advisor for EU AI Act Compliance. Unlock the full potential of artificial intelligence while ensuring compliance with the EU AI Act by partnering with AI & Partners, a leading professional services firm. We specialise in providing comprehensive and tailored software solutions for companies subject to the EU AI Act, guiding them through the intricacies of regulatory requirements and enabling responsible and accountable AI practices. At AI & Partners, we understand the challenges and opportunities that the EU AI Act presents for organisations leveraging AI technologies. Our team of seasoned experts combines in-depth knowledge of AI systems, regulatory frameworks, and industry specific requirements to deliver strategic guidance and practical solutions that align with your business objectives.

To find out how we can help you, email contact@ai-and-partners.com or visit https://www.ai-and-partners.com.

Contacts

Sean Donald John Musch, CEO, s.musch@ai-and-partners.com

Michael Charles Borrelli, Director, m.borrelli@ai-and-partners.com

Authors

Sean Donald John Musch, CEO

Michael Charles Borrelli, Director

Acknowledgements

Corporate Partners

We are grateful to our network of **corporate partners** for their invaluable contributions:

Individual Partners

We are also grateful to our network of **individual supporters** for their invaluable contributions:

Abhishek Gautam, AI Researcher || Project Assistant at IIT Dhanbad.

Doug Hohulin, coauthor of "Tech Power Healing - The Future of Medicine in the AI Age", is working on AI projects "When the AI System Has to Be Right": AI in Healthcare, Automated/Connected Vehicle, Governance/Policy and Energy. He worked for 33 years at Nokia/Motorola in Strategy, Business Development, Account Management and Engineering roles. He is focused on Responsible AI (Ethics, Governance, Policy, Regulation, Safety) and how AI can be used to help create a safer health system - reduce medical errors (3M deaths/year in the world), more patient adherence to the right treatment plan. He was a member for 4 years of the 5G Automotive Association working on the development of AV/CV. He is the VRARA Chapter President - KC / Generative AI Committee Chair. He is on the KU School of Nursing Advisory Board and supported various XR projects and a KC Digital Drive Member focusing on: Healthcare Innovation and provis consulting, training/workshops on Generative AI and other technologies.

Dr. Harvey Castro, Dr. Harvey Castro is an ER physician, AI futurist, and author specializing in AI's impact on healthcare. With over 20 years of medical experience and a strong background in AI, he bridges the gap between technology and patient care to enhance healthcare outcomes and safety.

Jerome Silber, Senior Fellow, a seasoned attorney with extensive experience as an external lawyer and in-house counsel. He specializes in negotiating commercial and technology agreements. Jerry has also taught communications law and ethics at various local colleges. He has been an active member of World Commerce & Contracting (including several terms on the Advisory Council), leading to his "Fellow" status award earlier this year. His commitment to pro bono work has earned him numerous awards. Jerry holds a bachelor's degree in Communications, a master's degree in Journalism, and a law degree. Since retiring from Verizon in March, Jerry has advised several start-ups on commercial and regulatory matters and consulted with Frontier Communications. He holds an advisory role at AI 2030 and has begun a Monthly AI Litigation Review.

Jonathan Gonzalez, Jonathan has 17 years' experience working with multilateral organisations (World Bank, Asian Development Bank, OECD, UNESCO, APEC, and ASEAN) and leading technology companies to develop whitepapers and reports on a wide range of policy issues, including the digital transformation of governments and the impact of AI in Asia's digital economies.

Kate Shcheglova-Goldfinch, Kate is on the list of Top3 UK Banker of the Year (Women in Finance Awards, 2023), TOP10 Policy makers and regulatory experts (Women UK Powerlist'23 by Innovate Finance), UN Women UK'24 Delegate developing tech diversity. Last four years Kate led the fintech and regulatory stream as an external Senior PM at EBRD, serving local central banks, creating regulatory acts and deploying regulatory sandbox legal frameworks (first global regulatory sandbox deployed under the war conditions for Ukrainian central bank, went live in March 23).

Matthew Gardiner, Matthew advises leaders of financial institutions, FTSE 100 and frontier tech companies, regulators and policy makers around the globe on developing and deploying GenerativeAI, web3 and frontier tech. Matthew has helped multiple companies swiftly move ahead of the curve and define their futures with emergent architectures.

Melanie Dow, Melanie has a plethora of experience in publishing, working at the intersection of blockchain and AI, she has her finger on the pulse of the latest technology.

Michael Boevink, Michael Boevink has more than 20 years management experience in the fintech and banking industry and is founder of his own investment company Boevink Group. Mr. Boevink specialises in capital raising, scaling and executing go-to-market strategies and business development growth in global markets and is engaged in companies as Raimac Financial Technology - Raimac.io - a programmable payment solution. He holds an MBA from the University of Bradford.

Parul Benien, Parul Benien is people-centric innovator, driving change and achieving results. Her expertise spans across business development, angel investing, venture capital, valorization, licensing, and the creation of spin-offs. She has a focus on data, AI and its intersection with diagnostics, global health, cancer research, drug delivery, and cell biology.

Paul Levy, Paul Levy is a senior lecturer at the University of Brighton, a teaching associate at Warwick Business School, author of the books Digital Inferno and The Poetry of Change.

Thea Montgomerie Anderson, Thea is the visionary co- founder of AI Spy Media, a dynamic brand encompassing a magazine, podcast, Substack, networking events, and media production services. AI Spy Media is dedicated to celebrating the positive side of AI and entrepreneurship, bringing the stories of founders, investors, and AI tech communities to life. Through these platforms, Thea aims to build a community that inspires and educates, showcasing the transformative potential of artificial intelligence.

Important notice

Opinions in this document reflect the opinions of the authors, and are not intended to be relied upon. The authors do not accept any responsibility for any reliance placed on this document. It is important to obtain professional guidance as appropriate when seeking to deal with the matters raised in this report.

AI & Partners B.V. is the Dutch headquarters of AI & Partners, a global professional services firm. Please see https://www.ai-and-partners.com/ to learn more about us.

© 2024 AI & Partners B.V. All rights reserved.

Designed and produced by AI & Partners B.V