Trustworthy AI:
The fundamentals

THALES
Building a future we can all trust

Editorial

In a constantly evolving world, an organization's ability to responsibly harness artificial intelligence (AI) has become a societal imperative.

Artificial intelligence is deeply transforming economies and ways of life, and it is our collective responsibility to ensure that this technological revolution is guided by ethical, transparent, and sustainable principles.

At Thales, we uphold the highest standards of responsible digital practices while supporting clients whose missions demand the highest standards of performance, resilience, and security. AI is at the heart of our strategy, as we design sustainable, sovereign, and trustworthy solutions.

This document presents a vision, commitments, and cutting-edge expertise in the field of digital services incorporating AI-based components. It encourages a rethinking of artificial intelligence as a technology serving humanity, the environment, and digital sovereignty.

At Thales, trusted AI is conceived as both a framework and a compass to address today's challenges, while creating **lasting and trustworthy value** for tomorrow. By collaborating with experts, specialized working groups, and national as well as European partners, Thales has positioned itself as a key player in Trustworthy AI that serves humanity. By going beyond the regulatory framework, our trusted AI approach acts as a lever to enhance business processes by defining shared responsibility among stakeholders, while simultaneously driving increased performance in our solutions and services.

THALES
Building a future we can all trust

Trustworthy AI

Kalina Genet, Edouard Roger, Clément Engel

June 2026

Foreword

Navigating the Trustworthy AI ecosystem

Artificial Intelligence (AI) is rapidly reshaping societies and economies at an unprecedented pace. As AI adoption accelerates, establishing **Trustworthy AI** has emerged as a pivotal challenge to ensure that technological progress is harmonized with responsibility, innovation, and sustainability. This article seeks to address the intricate landscape of trustworthy AI, providing a **conceptual framework for understanding the current state of the art** in France and across Europe, followed by an **illustrative example of practical implementation.**

The proposed framework is intended to be **interdisciplinary and readily accessible**, targeting stakeholders interested in embedding Trustworthy AI principles into their strategic agendas and day-to-day operations. It addresses a broad professional audience, spanning senior decision-makers, technical practitioners, cybersecurity specialists, Chief Data/AI Officers, legal professionals, and sustainability experts.

Our approach is deliberately open-ended: for each thematic area covered, more detailed analyses or sector-specific adaptations exist in various specialist studies, initiatives, and publications. It should be emphasized that the overview presented herein is inherently dynamic and is not intended to be comprehensive.

The AI system at the core of a responsible approach

To properly address the concept of Trustworthy AI, it is essential to clarify the definition of "AI" adopted in this position paper.

We adhere to the definition proposed in the European AI Act (Article 3.1)[1], which specifies an **AI system** as "*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.*"

Additionally, our approach encompasses general-purpose AI models as defined in the AI Act (Article 3.63): "*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.*"

[1] European Union, Artificial Intelligence Regulation (AI Act), " Regulation (EU) 2024/1689 " (2024)

Our systemic approach is therefore designed to address the **entire lifecycle of an AI system**, which combines a model with a given operational purpose, from its initial design through to deployment and ongoing maintenance. AI is not an end in itself but rather a tool in service of a defined need.

Responsible consideration of AI systems raises significant challenges for corporate strategy, notably the complexity of explicitly linking sustainability, competitiveness, sovereignty, and profitability.

Document structure

We seek to highlight the diversity of expertise required for the development of Trustworthy AI—including legal, environmental, cybersecurity, and technical domains—and to reference the major works addressing this complex and cross-cutting topic. Accordingly, we cite works and sources that range from foundational texts on the ethical governance of AI to technical specifications developed to mitigate risks arising from AI deployment.

Three major aspects are addressed in this document:

1. **Fundamental principles:** This section aims to define Trustworthy AI and to elaborate on the four key pillars underpinning it: **validity, security, transparency** and **responsability**. These principles result from an in-depth analysis of available documentation and reference frameworks on the market as well as Thales' own research.
2. **Regulatory landscape and reference works:** Here, we present a non-exhaustive overview of the major initiatives and key regulations regarding Trustworthy AI, with a particular focus on European and French legislative frameworks, as well as prominent organizations and stakeholders in the field. It should be noted that a number of existing regulations—including those related to personal data (such as the GDPR) or to cybersecurity (such as the Cyber Resilience Act)—apply indirectly to AI.
3. **Illustrating the implementation of trustworthy AI:** Finally, we aim to demonstrate how Trustworthy AI can be applied within a company such as Thales. This part serves to complement the first two sections, which address trustworthy AI from a global and holistic perspective. The implementation of Trustworthy AI may take multiple forms and initiatives with varying timelines.

All referenced sources mentioned in this overview are available via clickable hyperlinks at the end of the document.

1. Introduction

Artificial intelligence is transforming our world through its diverse applications and vast potential. However, this technological revolution should not come at the expense of our core values or the health of our planet. **Trustworthy AI** has consequently become a major priority in building a future where **innovation aligns with both responsibility and sustainability.**

Numerous concepts have emerged to characterize Trustworthy AI—including Responsible AI, Ethical AI, AI for Good, Green AI, and others—but what **concretely defines Trustworthy AI, and how can it be implemented in Europe?**

From a legal perspective, the regulatory framework governing the use of AI is constantly evolving. **Various legislative, regulatory, and standardization initiatives** at multiple levels have emerged in recent years, helping to outline the contours of Trustworthy AI:

> **International standards:** International bodies such as the OECD[2] and ISO[3] are working to develop non-binding standards and guidelines for Trustworthy AI. UNESCO[4] has also emphasized the importance of fostering AI that is ethical and respectful of human rights.

> **The European Union Artificial Intelligence Regulation (AI Act):** This landmark regulation, adopted in 2023, aims to establish a foundation of trust for the development and use of AI within the European Union. The Act sets out a differentiated regulatory framework based on the risks posed by AI systems, with specific obligations for high-risk systems. It helps delineate the notion of responsible AI, notably by imposing requirements for transparency and combating discriminatory biases. This constitutes an essential first step toward safeguarding security and fundamental rights.

> **National initiatives:** Governments—including France, with its national AI strategy—are also emphasizing the necessity of a regulatory framework for AI development. Many countries, such as the United States (notably California[5]), Canada, China, South Korea, Japan, and the United Kingdom, are advancing their own AI initiatives, often adopting markedly different approaches due to cultural and/or ideological differences. For example, the United States generally favors an "ultra-liberal" approach, while the European Union and South Korea tend toward more regulatory models.

> This evolving framework highlights the role of **research programs** in going beyond minimum legal requirements. Academic programs, industry consortia, national organizations, and think tanks are developing guidelines, reference frameworks, standards, methodologies, and supporting documentation to facilitate the interpretation of the law, anticipate future challenges, and promote proactive, responsible, and sustainable practices.

The primary drivers behind the regulation of Trustworthy AI are:

Security

AWARENESS OF RISKS

Ethics

THE NEED TO PROTECT FUNDAMENTAL RIGHTS

Transparency

THE DEMAND FOR TRANSPARENCY

[2] OECD, " Principles for Trustworthy AI " (2024)
[3] ISO/IEC 42001, " AI Management system " (2023)
[4] UNESCO, " Ethics of Artificial Intelligence " (2021)

[5] See article " Governor Newsom signs SB 53, advancing California's world-leading artificial intelligence industry " | Governor of California

2. Context and key issues

Beyond the ethical and legal considerations that typically emerge when discussing Trustworthy AI, another issue has become critical in light of recent findings on the energy consumption of generative AI models[6]: **sustainability**. AI indeed has a significant environmental footprint if not designed and deployed responsibly.

It is therefore essential to **place sustainability at the heart of AI-related concerns**. This involves reflecting on the environmental impact of AI systems—including not only algorithms but also the hardware infrastructures required for their operation (such as accelerators, data centers, and the use of rare metals); developing green AI solutions; and fostering a culture of responsible AI within enterprises and organizations.

Thus, Trustworthy AI represents a systemic approach that goes beyond **legal compliance or the mere minimization of potential risks.** It requires a **proactive commitment** to information, training, and engagement, incorporating ethical, social, and environmental principles.

But what does this **term truly encompass**? What reference frameworks **guide this approach**? And how are organizations striving **to embed it within their practices**?

Trustworthy AI is a collective challenge, and the following provides several key insights to address these questions, including:

- The **foundational principles of Trustworthy AI**, and the concepts that arise from these principles in a holistic framework.
- A **non-exhaustive overview of regulations and reference publications** available to address the requirements and challenges associated with each of these principles. All cited documentation is available at the end of the document.
- An illustration of **initiatives and the implementation** of Trustworthy AI within an organization such as Thales.

3. Defining Trustworthy AI

Understanding Trustworthy AI begins with clarifying the relevant terminology, which is essential for any initiative aiming to develop AI that is responsible, beneficial, and safe.

Surrounding the notion of Trustworthy AI is an **ecosystem of interconnected principles, concepts, and terms**, which can complicate interpretation. Terms such as Responsible AI, Ethical AI, and Trustworthy AI are sometimes used interchangeably, though each refers to distinct dimensions (see Figure 1).

- **Sustainable AI**: Sustainable AI aims to design AI as *"a strategic technology for accelerating ecological transition and better managing our planet's resources."*[7] In this sense, AI should be leveraged both to support the ecological transition (e.g., enhancing our understanding of climate dynamics, optimizing resource flows) and to improve its own environmental efficiency (e.g., energy optimization, eco-design). The term "AI for Green" has emerged to describe the former, while "frugal AI" or "Green AI" refers to the latter.
- **Ethical AI**: Ethical AI[8] focuses on moral principles, rooted in societal values which should guide the development, deployment and use of AI in a virtuous manner[9], specifically through the moral and legal dimensions of upholding fundamental rights and regulatory compliance. Closely linked to the aims of Sustainable AI, "AI for Good" designates the application of AI for positive societal impact in service of the common good.
- **Transparent AI**: AI transparency seeks to *"enable individuals to better understand how AI systems are created and how decisions are made,"*[10] by providing explanations that are both understandable and contextually appropriate. Explainability and knowledge of model functioning and algorithmic logic are fundamental to Trustworthy AI, as they strengthen trust in outputs, as well as the efficiency and fairness of decisions.
- **Secure and Safe AI**: Secure and Safe AI aims to ensure the robustness and resilience of systems against adverse conditions such as data leaks, model manipulation, spoofing, and cyberattacks. AI safety aims to address inherent issues and unintended consequences, while AI security focuses on protecting

[6] *See article "Intelligence artificielle en entreprise : quels impacts environnementaux ? "| Direction générale des Entreprises*
[7] *Source : DP_Coalition_mondiale_IAdurable.pdf | Sommet pour l'action sur l'IA*
[8] *Source : Definition " Ethical AI " - HLEG European Commission, " Ethics Guidelines for Trustworthy AI " (2019)*
[9] *Source : ISO, " ISO - Building a responsible AI: How to manage the AI ethics debate "*
[10] *Source : What Is AI Transparency? | IBM*

AI systems from external threats. A safe AI system is a hallmark of reliability and, in this regard, provides a foundational framework for trust in AI.

- **Responsible AI**: Responsible AI (RAI) is an umbrella term that refers to the choices when adopting AI related to the design, development, deployment, and oversight of artificial intelligence systems in ways that are aligned with legal and societal expectations while minimizing the risk of negative consequences. Responsible AI builds trust in AI solutions by considering the "*broader societal impact of AI systems and the measures required to align these technologies with stakeholder values, legal standards and ethical principles*"[11]. This notion refers to the dimension of operational accountability (governance, auditability, etc.) and aims to guarantee compliance with regulations (AI Compliance). For Thales, responsibility also means alignment with its Digital Ethics Charter[12]. It involves implementing governance frameworks, processes, roles, audits, and mechanisms of accountability to ensure that AI systems are developed and used in an ethical manner that benefits society, while minimizing risks[13]. The responsibility of an AI system primarily resides with its stakeholders, defined as "*all organizations and individuals involved in, or affected by, AI systems, directly or indirectly*" (including designers, integrators, users, regulators, and others)[14].
- **Trustworthy AI**: The European Commission popularized the concept of Trustworthy AI, particularly through the work of the High-Level Expert Group on Artificial Intelligence (HLEG AI)[15]. This concept aims to translate ethical and sustainable principles into concrete, measurable requirements. Trustworthy AI, according to the European Commission, is AI that is safe, transparent, ethical, unbiased and under human control[16]. It integrates technical aspects such as robustness, reliability, and security, along with the legal dimensions of ethical AI, including data protection (e.g., GDPR) and compliance with the AI Regulation.
  Thales is aligned with this vision—through its TRUE AI[17] approach—by designing Trustworthy AI that incorporates all key dimensions of AI (ethical, sustainable, secure, transparent, responsible, etc.) and meeting strict sovereignty requirements in service of strategic and critical sectors.

A Trustworthy AI ensures systems **validity, explainability, security, and responsibility.**

Validity is the foremost fundamental principle for any AI deployment. Validity ensures that an AI system does what it is supposed to do, does all that it is supposed to do, and only what it is supposed to do, by confirming—with objective evidence—that "*the requirements for a specific intended use or application have been fulfilled.*"[18]

On this basis, we propose to present these various dimensions of Trustworthy AI below, by positioning **the concepts and key principles** previously described and frequently referenced in the literature.

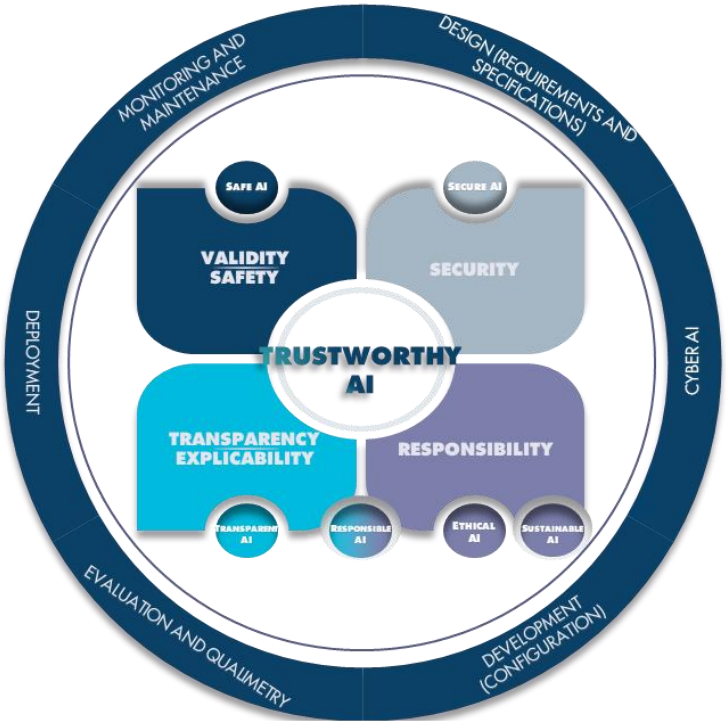


Figure 1: Trustworthy AI and Its alignment with the AI Lifecycle.
The principles of validity, security, transparency, and responsibility form the foundation upon which the governmental, technical and operational aspects of the trustworthy AI approach are built.

[11] *Source: What is responsible AI? | IBM*
[12] *Presentation of the Charter in the section 5.1.1 – Frameworks fondamentaux*
[13] *Scope « Accountability » covered by " AI principles | OECD " adopted in May 2019 by the OECD*
[14] *Reference to the concept of "stakeholders "and " Accountability " OECD, « Recommendation of the Council on Artificial Intelligence » (2019)*
[15] *HLEG European Commission, " Ethics Guidelines for Trustworthy AI " (2019)*
[16] *Source: European Commission, " Excellence and trust in artificial intelligence - European Commission "*
[17] *Transparent, Reliable, Understandable, Ethical AI décrit dans le Position Paper Thales « Cycle de vie d'ingénierie IA fiable de bout en bout pour les systèmes critiques »*
[18] *Source: ISO/IEC 22989:2022, " Artificial intelligence concepts and terminology " (2022)*

4. Trustworthy AI components

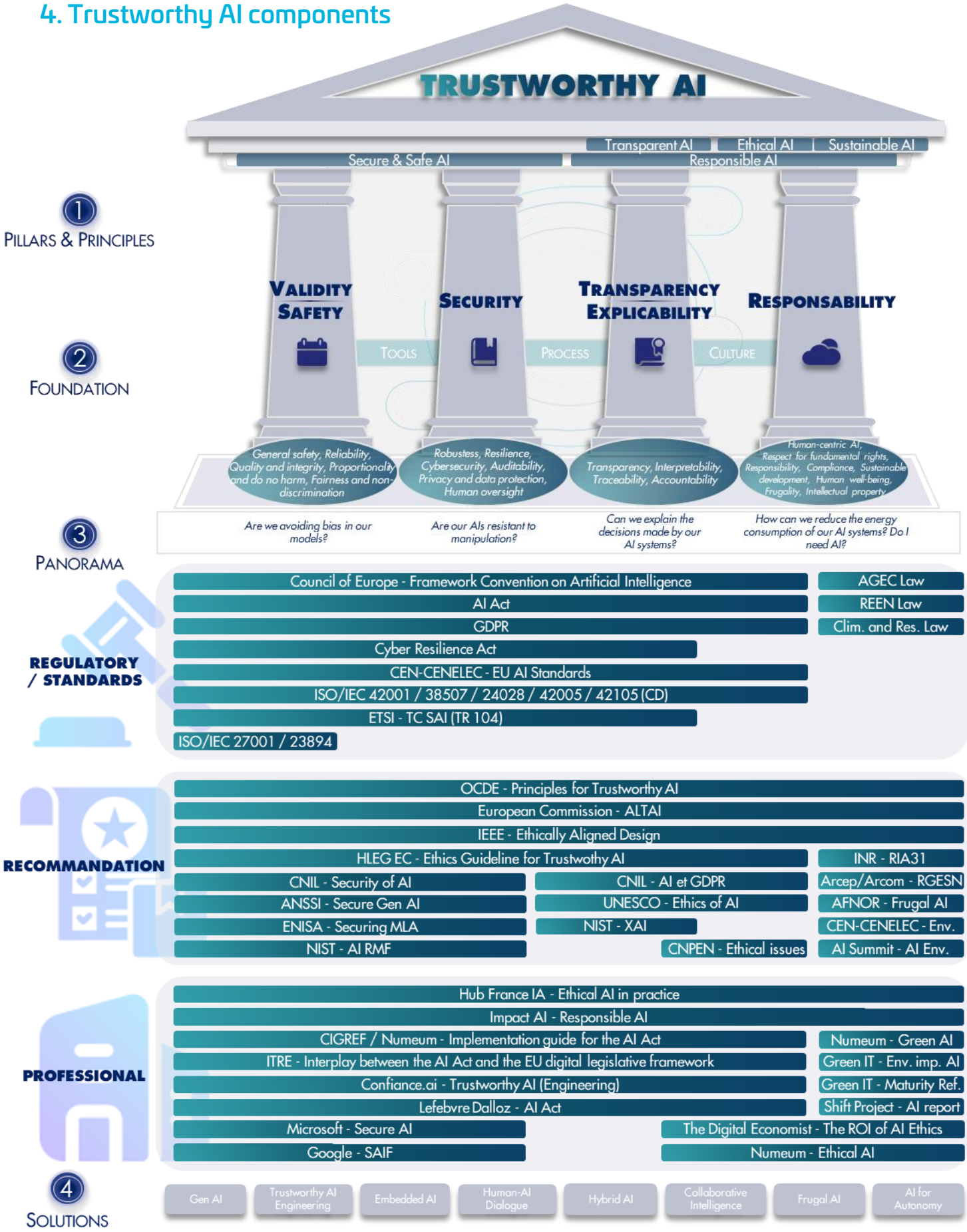


Figure 2: The temple of Trustworthy AI

4.1 PILLARS AND PRINCIPLES OF TRUSTWORTHY AI[19]

VALIDITY & SAFETY

This pillar encompasses aspects related to the evaluation and certification of reliability, performance, and ethics (particularly regarding bias), ensuring the validity of an AI system, as well as safety aspects aimed at preventing accidents, internal malfunctions, or harmful consequences associated with its use.

- **General safety plan**: Ensures that the system will do "what it is supposed to do without harming living beings or the environment. This includes the minimization of unintended consequences and errors." To this end, processes must be implemented to clarify and assess the potential risks associated with the use of AI systems. "*The level of safety measures required depends on the magnitude of the risk posed by an AI system, which in turn depends on the system's capabilities. Where it can be foreseen that the development process or the system itself will pose particularly high risks, it is crucial for safety measures to be developed and tested proactively." [EC]*
- **Reliability:** "*The property of consistent intended behavior and results*" *[ISO/IEC]. "A reliable AI system is one that works properly with a range of inputs and in a range of situations,"* enabling examination and prevention of unforeseen harm. [EC]
- **Quality and Integrity:** Ensures that data, models, inputs, and outputs are not altered (no inappropriate modifications or malicious data) and are of high quality. *[EC]*
- **Proportionality and Do No Harm:** *"The use of AI systems must not go beyond what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such uses." [UNESCO]*
- **Fairness and non-discrimination:** *"AI actors should promote social justice, fairness, and non-discrimination while taking an inclusive approach to ensure AI's benefits are accessible to all." [UNESCO]* The principles of fairness and non-discrimination aim to counteract bias right from the AI system's design stage (e.g., by ensuring the diversity of training data, monitoring of these data and algorithms, and transparency of AI system decision-making...).

SECURITY

This pillar covers measures designed to protect users, systems and data from risks or threats posed by cyberattacks and external threats.

- **Robustness:** The ability of an AI system *"to maintain its level of performance under all circumstances" [ISO/IEC].* The robustness of an AI system is linked to damage prevention, encompassing robustness *"from a technical perspective ([…] in a given context, such as the application domain or life cycle phase) and from a social perspective (in due consideration of the context and environment in which the system operates),"* so that AI systems *"behave as intended while minimizing unintentional and unexpected harm, and preventing unacceptable damage." [EC]*
- **Resilience:** "*The ability of a system to recover operational condition quickly following an incident*" *[ISO/IEC].* Resilience against attacks on AI systems involves protection against vulnerabilities (in data, models, or infrastructures) that may allow exploitation *"by malicious intention or by exposure to unexpected situations." [EC]*
- **Cybersecurity:** The set of defensive measures that ensure the integrity, confidentiality, availability, and traceability of an AI system against cyber threats. These measures cover training data, system development, and system operation, as well as cross-functional measures. *[CNIL]*
- **Auditability:** *Involves "enabling the assessment of algorithms, data, and design processes [...] Evaluation by internal and external auditors, and the availability of such evaluation reports, can contribute to the trustworthiness of the technology." [EC]*
- **Privacy and data protection:** Ensures *"privacy and data protection throughout a system's entire lifecycle. This includes the information initially provided by the user, as well as information generated about the user over the course of their interaction with the system". [EC]*
- **Human oversight:** Ensures that an AI system does not make critical decisions autonomously and remains under human control or does not cause adverse effects. *"Oversight may be achieved through governance mechanisms such as a 'human-in-the-loop' (HITL), 'human-on-the-loop' (HOTL), or 'human-in-command' (HIC) approach." [EC]*

[19] ***Sources Validity & Safety***: European Commission (EC), "Ethics Guidelines for Trustworthy AI" (2019); ISO/IEC 22989:2022, "Artificial intelligence concepts and terminology" (2022); UNESCO, "Ethics of Artificial Intelligence - AI | UNESCO";
Sources Security: ISO/IEC 22989:2022, "Artificial intelligence concepts and terminology" (2022); European Commission (EC), "Ethics Guidelines for Trustworthy AI" (2019); CNIL « IA : Garantir la sécurité du développement d'un système d'IA | CNIL »
Sources Transparency: OECD, "Recommendation of the Council on Artificial Intelligence" (2019); ISO/IEC 22989:2022, "Artificial intelligence concepts and terminology" (2022); IBM, "What Is AI Interpretability? | IBM " (2024); IEEE, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems" (2019)
Sources Responsibility: European Commission (EC), "Ethics Guidelines for Trustworthy AI" (2019); OECD, "Recommendation of the Council on Artificial Intelligence" (2019); UNESCO, "Ethics of Artificial Intelligence - AI | UNESCO"; IEEE, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems" (2019); European Union, "AI Act " (2024) ; WIPO, "WIPO Conversation – AI Inventions" (2024); CNIL, "Déclaration commune sur la gouvernance fiable des données pour l'IA" (2025) ; IBM, "AI Compliance: What It Is, Why It Matters and How to Get Started | IBM"UNESCO, "Artificial Intelligence and Sustainable Development" (2019); AFNOR, "IA frugale" (2024); UNESCO, "Ethics of Artificial Intelligence - AI | UNESCO"

TRANSPARENCY & EXPLAINABILITY

This pillar ensures that AI decisions are understandable, accessible, and that it is possible to trace processes for clear explanations.

- **Transparency:** Aims to make AI decisions and processes explainable so that individuals affected by an AI system can comprehend and grasp the outcomes *[OECD]*. It is "*the property of an AI system that appropriate information about the system is made available to relevant stakeholders.*" *[ISO/IEC]*
- **Interpretability:** Whereas explainability aims to "*provide reasons for the AI model's outputs,*" interpretability focuses on "*understanding the inner workings of an AI model*" to better comprehend, clarify, and make accessible their decision-making processes (architecture, features, biases, etc.). *[IBM]*
- **Traceability:** Complementing security and safety, the traceability of "*datasets, processes, and decisions made during the AI system lifecycle, to enables analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art,*" through detailed technical tracking and documentation. *[OECD]*
- **Accountability:** Transparency ensures accountability, which entails that the AI system must be "*created and operated to provide an unambiguous rationale for all decisions made.*" This concerns the obligation to technically account, through monitoring mechanism, for the effects caused by AI, by adhering to regulatory requirements and making the underlying logic of the AI system intelligible, ensuring it provides an unambiguous justification for all decisions taken. *[IEEE]* Accountability involves demonstrating the process behind a decision (data sources, algorithm implemented, and validation procedures), while responsibility refers to identifying the individual or entity liable for the outcome.

RESPONSABILITY

This final pillar encompasses principles regarding the responsibility of stakeholders, particularly with regard to compliance, ethics, and sustainability in AI. The goal is to promote AI that respects moral and human values, taking into account environmental and societal impacts, in support of human well-being and sustainable development.

- **Human-centric AI:** Approach that emphasizes respect for the primacy of "*the rule of law, human rights, and democratic values, throughout the AI system lifecycle*", including freedom, human dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, and social justice. *[EC, OECD]*
- **Respect for fundamental rights:** "*Rights enshrined in the EU Treaties, the EU Charter and international human rights law.1*" (with reference to dignity, freedoms, equality and solidarity, citizens' rights, and justice) that provide "the most promising foundations *for identifying abstract ethical principles and values, which can be operationalized in the context of AI.*" *[EC]*
- **Responsibility of actors:** Establishing (moral and legal) responsibility for all stakeholders in the AI value chain regarding the proper operation of AI systems throughout their lifecycle, in accordance with legal requirements and the state of art. *[OECD, UNESCO, IEEE, AI Act]*
- **Regulatory compliance (legal framework, standards, and norms):** Compliance in AI refers to "*decisions and practices that enable businesses to stay in line with the laws and regulations that govern the use of AI systems. These standards include laws, regulations and internal policies designed to help ensure that organizations develop AI models and their algorithms responsibly.*" *[IBM]*
- **Sustainable development:** Promotes research, multi-stakeholder collaboration, equitable access to AI, and awareness of ethics and critical thinking to prevent negative effects on societies. This includes facilitating access to inclusive data infrastructures (to avoid bias) and sustainable computing resources to align AI growth with global climate goals (notably the 17 SDGs adopted by the UN in 2015). AI technologies should be assessed against their impacts on sustainability. *[UNESCO, OECD]*
- **Human well-being:** Promotes *"the quality of life of individuals and communities"* through the use of AI. *[OECD]*
- **Frugality:** Overall reduction, from the design stage, "*of material and energy requirements and associated environmental impacts by redefining use cases or performance requirements, or by redirecting the needs from the AI system's producer to the service provider concerned,*" in order to limit its ecological footprint. *[AFNOR]*
- **Intellectual property:** Issues relating to the protection of training data (works protected by copyright), respect for and recognition of the intellectual property of inventions created by AI (data, algorithms, models), incorporating the use of AI or produced with the help of AI. *[WIPO]* Data protection principles must be integrated from the design stage. *[CNIL]*

4.2 ADOPTION FOUNDATION

While the four pillars constitute the essential principles of trustworthy AI, their consolidation intrinsically depends on a strong organizational foundation built upon: **appropriate tools, well-defined processes, and a committed corporate culture.**

In general, the design and deployment of AI must be accompanied by in-depth reflection on its social acceptability. Social acceptability, in the context of AI, refers to the degree to which the technology is perceived as **appropriate, fair, and beneficial** by individuals and society as a whole. UNESCO, for instance, in its Recommendation on the Ethics of Artificial Intelligence (2021), highlights the importance of inclusive stakeholder participation to ensure that AI systems are "acceptable to society." Other research, notably from the European Commission in its White Paper on Artificial Intelligence (2020), also emphasizes the importance of public trust and buy-in for the successful development of AI.

Beyond technical feasibility, adopting **a trustworthy AI approach is based above all on questioning the true necessity and impact of AI.** This means considering, for each new project, whether deploying AI is genuinely required to solve the problem at hand.

This topic becomes even more important when viewed through the lens of the **Jevons paradox**, which presents a major challenge. This paradox states that an increase in efficiency in the use of a resource may paradoxically lead to an overall increase in its total consumption. Applied to AI, it means that if AI renders certain processes more efficient, it may simultaneously encourage broader and potentially excessive use, raising questions regarding its **genuine necessity and overall impact on the environment,** employment, and data consumption. The question of social acceptability and the deeper reasons behind our technological choices therefore becomes crucial.

Such questioning, along with the initial cost of trustworthy AI—including consideration of ethical and environmental criteria at the design stage, rigorous model validation, transparency, and governance—may seem high and weigh on competitiveness. However, this cost must be viewed as a strategic investment, especially for critical systems, to ensure **long-term viability in a technological ecosystem increasingly focused on responsibility and ethics.**

Tools serving people and promote responsibility

> A trustworthy AI approach becomes significantly more effective when supported by a suitable technological toolbox. This includes tools to protect and secure algorithms, such as platforms ensuring traceability of data and models; solutions for AI explainability, which facilitate understanding of algorithmic decisions; life cycle analysis (LCA) methods for eco-design; and auditing instruments to detect and correct potential biases. These tools are not merely add-ons—they are catalysts that transform intentions into concrete actions, providing the technical means to measure, control, and enhance the responsibility over AI systems.

Integrated processes

> Trustworthy AI must be embedded into the very fabric of development and deployment processes of each AI solution, as well as its governance. This involves establishing specific roles, regularly reviewing regulatory compliance, upholding fundamental rights, incorporating ethical and security considerations from the design stage, integrating systematic impact assessments, and defining feedback loops for continuous improvement. By incorporating these requirements into the key stages of the AI lifecycle, responsibility and accountability is ensured not as an option, but as an operational requirement at every level.

A responsible corporate culture

> Finally, the deepest transformation occurs at the human level. Building trustworthy AI is above all a matter of a corporate culture focused on governance, role definition, and social acceptability. This involves raising awareness, training, and empowering all teams—from developers to decision-makers—on the ethical, societal, and environmental issues surrounding AI. By fostering a culture centered on diligence, accountability, and critical curiosity, individuals are encouraged to question, be proactive, anticipate risks, and actively engage in the design of AI systems that not only support the organization's goals but also promote more virtuous uses of AI. A strong corporate culture is the glue binding processes and tools, transforming abstract principles into daily and sustainable practices.

Trustworthy AI is an essential approach to ensure that AI is used in a way that **benefits everyone, while remaining at the service of humanity.**

4.3 OVERVIEW OF THE LITERATURE

A **wide range of stakeholders** contribute to the ongoing discussions and regulations surrounding the various principles of trustworthy AI: research centers, business associations, professional communities, administrative authorities, and international or governmental organizations. For each pillar, as well as their associated concepts, **reference frameworks, guidelines, and practical or methodological guides** can be found to best address the challenges at hand.

This document presents a selection of the identified documentation across **three levels:**

There are various frameworks—including legal bases, regulations, standards, or norms—that address regulatory requirements or ensure compliance with recommendations issued by standardization bodies. Regulation and standardization are generally regarded as complementary aspects.

Work carried out by administrative authorities, ministerial missions, international or governmental organizations takes the form of reference frameworks, guides, practical sheets, and recommendations. The main objective is usually to provide tools, frameworks, and best practices to facilitate the application of the law or address the challenges posed by trustworthy AI.

Work undertaken by professional or business associations, think tanks, as well as professional or corporate communities, typically consists of guides, methodological documents, or research addressing the themes of trustworthy AI. These documents are often focused on the practical applicability of a law or principle, the specificities of a given sector, or the use of particular tools.

All referenced documentation is available at the end of this document (Section 7: Resources). For each topic discussed, more in-depth or sector-specific materials can be found within the working groups, organizations, or institutions mentioned. Readers are therefore encouraged to explore the **additional resources** offered by these key stakeholders according to their particular needs or interests.

4.4 SOLUTIONS

The deployment of Trustworthy AI is a **multidimensional challenge**, encompassing aspects from governance to the selection of appropriate technology. The overview presented highlights governance, regulatory, and certain technical dimensions related to security, but it is important to emphasize the role of technology in order to obtain **a comprehensive view of AI system deployment**.

The careful selection of learning-based AI technologies is central to a Trustworthy AI approach, insofar as it must address the **needs and constraints specific to particular sectors or professional domains.**

AI technologies address particular facets of Trustworthy AI, and within Thales we have identified differentiating technologies specifically aimed at **the needs of critical systems**, which must incorporate security and safety constraints commensurate with their level of criticality.

These technologies are described in the article "Artificial Intelligence for Critical Systems" (2024) [20] (see Figure 3) and are intended to provide **concrete responses to the challenges posed by AI.**

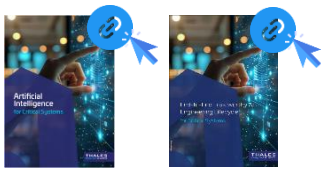
To operationalize responsible and trustworthy AI, technologies such as hybrid AI enable the combination of the strengths of connectionist methods with the logic and reasoning of symbolic AI to ensure transparency, interpretability and robustness; frugal AI, an approach that seeks to minimize environmental footprint and deployment costs; and self-explainable AI, which is essential for building trust by making AI system decisions understandable and justifiable to humans.

The implementation of Trustworthy AI can therefore take **multiple forms and unfold over different timeframes** depending on needs, stakes, organizational structures, sectors and contexts.

Thales' conviction is that, when properly controlled, AI can provide valuable support to humans, enabling them to make better and faster decisions. In this context, **AI should be viewed as a complement to human intelligence.**

Thales focuses on continuously renewing its solutions to achieve ever higher levels of performance, security and sustainability. This approach is embodied in part through the deployment of **Trustworthy AI and a range of initiatives to design reliability into AI from the outset from data and algorithms through to validation.**

Learn more :



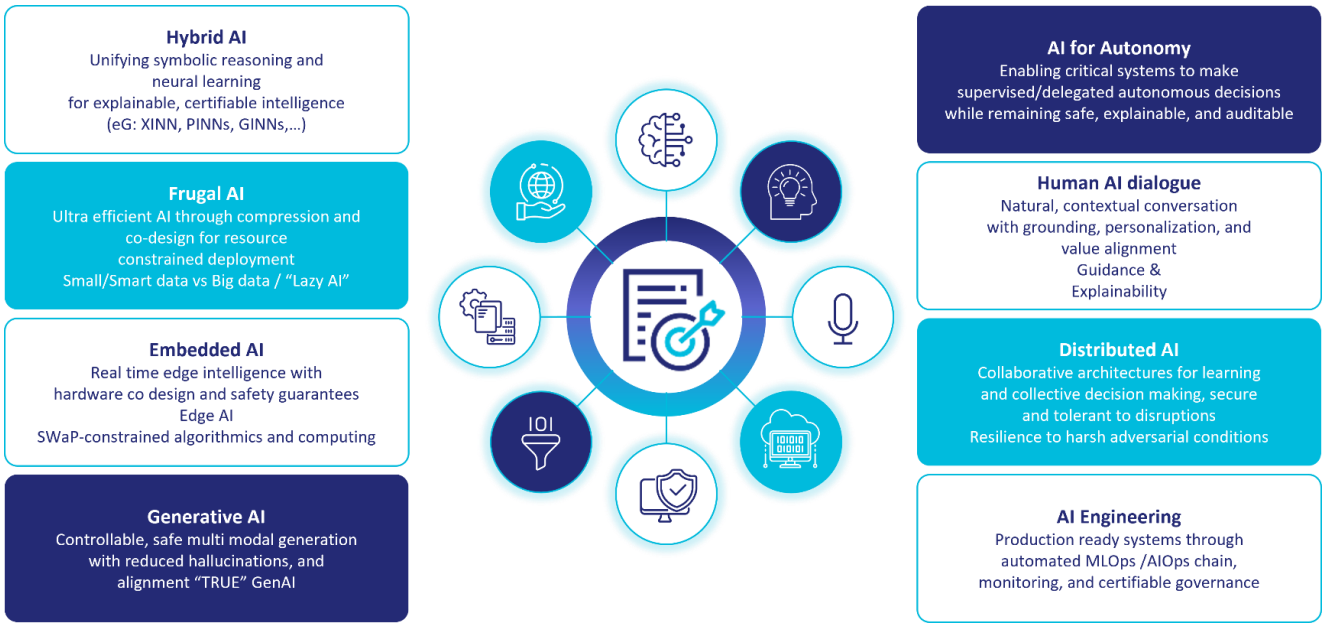


Figure 3: Differentiating AI technologies

[20] *Juliette MATTIOLI et Christophe MEYER, "AI for critical systems", (2024)*

5. Thales Initiatives to Implement Trustworthy AI

Within Thales, Trustworthy AI represents a **deep commitment** that goes beyond regulatory compliance and has guided the company's innovation strategy for several years. Trustworthy AI encompasses very specific aspects (such as security, transparency, ethics, sustainability…) that have been implemented at different times, drawing on diverse areas of expertise and coordinated appropriately. Beyond the expertise of its actors and their coordination, the deployment of Trustworthy AI depends on many factors, such as convictions about digital technology and AI, commercial and development strategy, the nature of products and services, internal organization, sector of activity, applicable regulation, etc.

Within the Thales Group, reflections on ethics and digital governance have been embedded since the earliest uses of technologies and AI in its projects. In 2019, the foundations of Trustworthy AI within the Group were laid and deployed with the **TRUE AI initiative**, followed in 2022 by the **Digital Ethics Charter**, which provides the guiding principles for responsible and trustworthy digital practices and applies across all Group activities, in alignment with our company purpose. The TRUE approach is structured around four pillars to implement Trustworthy AI: Transparency, reliability, explainability and ethics. This framework aims to ensure the integration, from the design stage, of **strong requirements for transparency, understanding of AI model functioning and respect for ethical principles.**

This responsibility is expressed across all its dimensions (technological, societal and environmental) to ensure that designed AI is not only performant and reliable, but also sustainable in the long term.

This responsibility is expressed across all its dimensions (technological, societal and environmental) to ensure that designed AI is not only performant and reliable, but also sustainable over the long term.

Implementing a Trustworthy AI approach requires the **mobilization of a variety of actors** including technical, legal, cyber and product/service specialists deployed across multiple levels including tools, processes and corporate culture. To steer this approach and ensure AI governance, it is essential to rely on these actors with a cross-functional unit that **ensures AI systems' compliance with European and international standards.**

Trustworthy AI implies numerous principles as described previously, such as security, safety and transparency. **With the ambition to develop Trustworthy AI**, as early as 2018 Thales contributed to the foundational work of the **DEEL** (DEpendable Explainable Learning) project, which aimed to develop technological building blocks for reliable, robust, explainable and certifiable artificial intelligence applied to critical systems. These scientific efforts, which brought together French and Quebecois researchers, contribute to the Trustworthy AI approach by addressing challenges related to robustness, explainability, uncertainty quantification, detection of distributional errors and industrial biases. Following this direction, Thales became in 2021 one of the founding members of the **Confiance.ai** program, which aims to make France "one of the leaders in Industrial and responsible AI by developing a sovereign, open, interoperable and sustainable methodological and technological framework, thereby accelerating the integration of trustworthy AI in strategic industries."[21] Confiance.ai is a community dedicated to accelerating the deployment of responsible and trustworthy AI in industrial systems, bringing together major industrial and academic actors. Its objective is to develop methods, tools and standards to support research and engineering in the development of robust, secure, explainable and controlled AI systems throughout their life cycle via a use-case driven approach. Following Confiance.ai in 2024, **the European Trustworthy AI Association** was created to continue animating the Trustworthy AI engineering community and to preserve the results, methods and tools. Directly following on from these efforts, the **CSIA ("Confiance dans les systèmes d'IA")** research program expands the methodological scope. Whereas **Confiance.ai** focused primarily on machine learning models, **CSIA** complements and adapts engineering methods to incorporate hybrid AI (combining learning and symbolic reasoning) as well as generative AI. The goal is to align architectures, validation processes and risk management frameworks with these new paradigms while preserving the requirements specific to critical systems.
Through these initiatives, Thales fully participates in a national, collaborative ecosystem dedicated to Trustworthy AI applied to critical systems, while taking into account the European regulatory framework of the AI Act.

In 2024 Thales presented its **acceleration mechanism for integrating artificial intelligence applied to critical systems**, particularly in the defence sector, while building on its technical expertise and mastery of the full range of key AI technologies. The creation of **cortAIx** aims to gather the Group's AI capabilities in research, sensors and critical

[21] *Confiance.ai : Home - Confiance.ai*

systems to ensure a robust and responsible application of AI. This acceleration program is applied in particular to the Defence sector, which requires consideration of strong and specific constraints, such as cybersecurity, embeddability and frugality, linked to critical environments[22].

The Thales Group thus positions itself among **European leaders in patent filings related to artificial intelligence** applied to critical systems and already integrates AI into more than a hundred of its products and services. While this often results in advances in security, transparency and ethics, Thales also places strong emphasis on environmental commitment, through participation in programs and the development of solutions to address, for example, **sustainable aviation** (Green aviation)[23], thereby reconciling technological innovation and sustainability in critical sectors such as aeronautics and space.

At the same time, involvement in European bodies and professional organizations also enables the Thales Group to contribute to the **development of reference frameworks adapted to critical systems and industrial realities**, while integrating security and sustainability issues at the heart of deliberations.

Within Thales, AI is conceived as a lever to meet today's challenges while creating enduring and **trustworthy value for tomorrow.**

5.1 Thales' Frameworks and Collaborative Dynamics

Artificial Intelligence at Thales encompasses a wide range of areas, from fundamental research to concrete applications serving both civil and defence needs. The development of Trustworthy AI relies on processes, tools, and a dedicated culture, which is reflected in core digital frameworks as well as specialized working groups.

5.1.1 – Core frameworks

Group-level frameworks establish a set of guidelines on digital trust and responsibility, which are essential for Thales to align its innovation priorities with its commitments to building a safer, greener, and more inclusive world. These frameworks reflect the development of numerous processes and ongoing considerations on ethics and digital governance, initiated from the first use of new technologies in Thales projects.

- *Digital Ethics Charter*

Context et objectives

To establish guidelines for responsible, secure, and sustainable digital practices, while addressing ethical, environmental, and societal challenges. The Charter is a public statement of Thales' commitment to using digital technology in a responsible, ethical, and environmentally respectful manner.

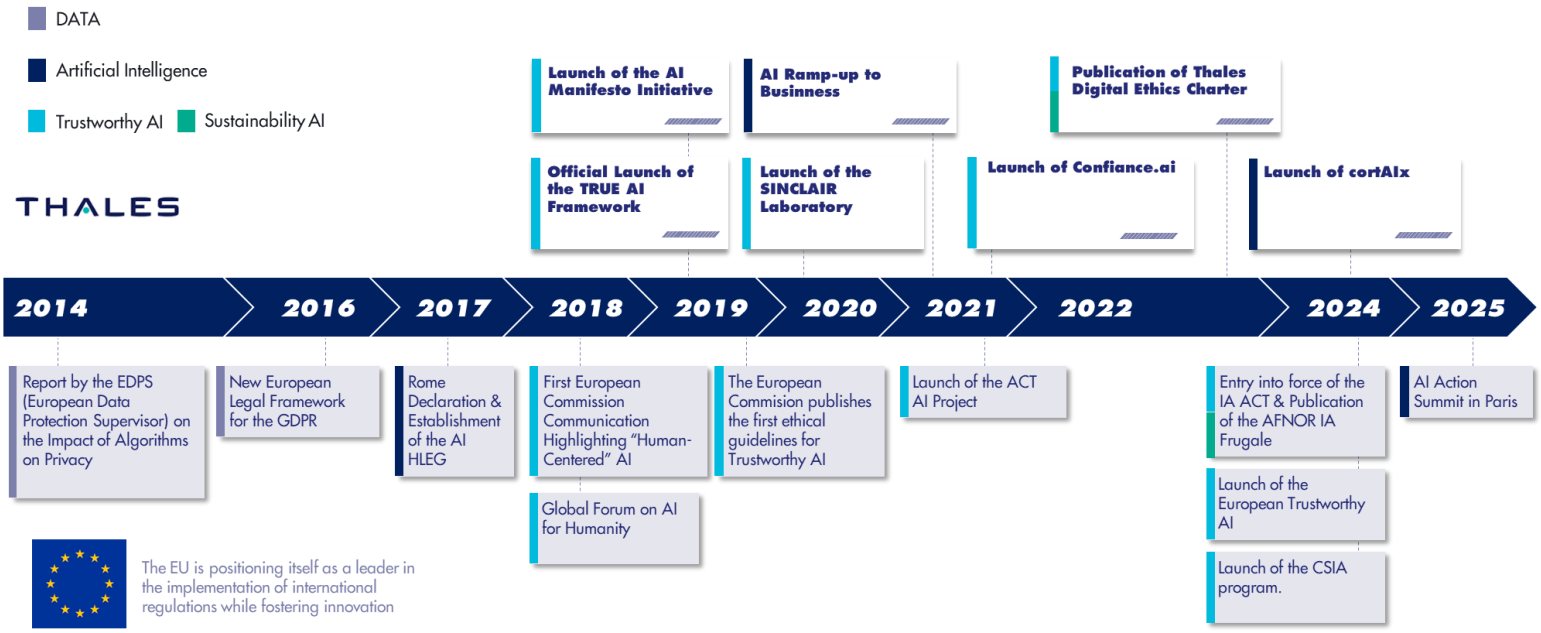


Figure 4: European AI Regulations and AI #Thales

[22] *See article : Thales accélère dans l'IA pour la défense | Thales Group*

[23] *See article : Les enjeux techniques et technologiques sont indissociables des enjeux environnementaux | Thales Group and this document*

Targets and Scope

This initiative applies to all Thales products, solutions, and services, and is intended for everyone: employees, customers, partners, and citizens concerned with digital issues.

Description

Thales defines ten commitments regarding digital trust and responsibility, structured around four priority areas:

- **Maintaining** human control and transparency of AI systems.
- **Strengthening** the safety and security of solutions.
- **Fostering** a more environmentally respectful world.
- **Placing** humans at the center of technologies to promote a more inclusive and equitable world.

Results

- Improved security and resilience of digital solutions.
- Reduced environmental impact.
- Enhanced protection of personal data.
- Contribution to fighting climate change.
- Adoption of ethical and responsible practices

Figure 5: The 10 Principles of Thales' Digital Ethics Charter

◇ TRUE AI (Transparent, Reliable, Understandable, Ethical)

Context et objectives

To support digital transformation and the growth of artificial intelligence, Thales has developed the TRUE AI approach (Transparent, Reliable, Understandable, Ethical), providing responsible products and services aligned with the Group's commitments to digital responsibility and ethics. The objective is to build greater trust among users and partners, which is an essential condition for adopting sensitive technologies.

Targets and Scope

- End users,
- Customers,
- Partners,
- Regulators.

With both national and international reach, TRUE AI is embedded in Thales's Digital Ethics Charter and complies with regulatory frameworks (e.g. GDPR, AI ACT).

Description

TRUE AI is based on four major principles of trustworthy artificial intelligence, applied to all Thales Group solutions:

- **Transparent**: Clear design and deployment guidelines, respect for confidentiality and consent.
- **Reliable:** Reliability assurance for "mission, safety, and business-critical systems".
- **Understandable**: Transparency regarding usage and results to ensure acceptability.
- **Ethical**: Regulatory compliance, adherence to standards, and the promotion of non-discrimination and equality.

Results

- **Establishing** a clear framework for the responsible use of AI and sensitive technologies.
- **Building** trust in and acceptance of Thales solutions.
- **Positioning** the group as a reliable and responsible partner.

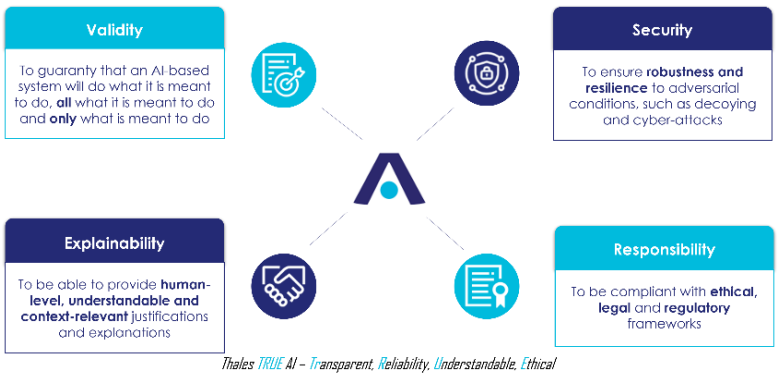


Figure 6: Thales TRUE AI – Transparent, Reliable, Understandable, Ethical

5.1.2 – Examples of national and European contributions and collaborations

Thales plays a driving role in the national and European AI ecosystem by contributing to ***working groups and collaborative projects****, in order to translate work into* ***methodologies or standards applied to industrial or civil challenges.*** *These initiatives, conducted in partnership with industry players, research laboratories, and institutions, are part of a dynamic of national and European cooperation aimed at strengthening technological sovereignty and guaranteeing the development of Trustworthy AI.*

◇ *Confiance.ai, Euro CSIA et European Trustworthy AI Association (#Safety #Security #Transparency #Responsability)*

Context and objectives

R&D program (2021–2024) stemming from the Grand Défi "Securing, certifying and making reliable AI-based systems", bringing together 13 French industrial and academic players including Thales as a founding member. This collective aims to develop a methodological and technological foundation enabling the industrialization of reliable and certifiable AI, in line with the AI Act, in order to address the challenges of robustness, explainability, and certification of critical systems.

Targets and scope

The program targets strategic industries such as aeronautics, defence, automotive, and energy, with direct application to concrete use cases.

Description

The work is structured around five research areas — reliability, explainability, data engineering, trust by design, and embedded AI — coordinated by IRT SystemX and deployed through industrial use cases involving Renault, Valeo, Thales, and other players.

Results

- Development of a tooled methodology and an open-source catalogue of components;
- First concrete applications, notably at Thales for the explainability of object detection algorithms.

Conclusion and outlook

The capitalization of Confiance.ai's work is formalized through the creation of the European Trustworthy AI Association and a CSIA working group, extending the program by broadening the methodology, in order to disseminate and sustain the results, contribute to standards, and prepare for upcoming developments (generative AI, cybersecurity, AI hybridisation).

◇ *WG 114 Artificial Intelligence (#Safety #Transparency)*

Context and objectives

Created by EUROCAE, WG-114 brings together industrial players such as Airbus, Thales, and Collins, as well as authorities including EASA and FAA, and research laboratories, in order to frame the use of AI in aviation — a sector where certification is complex due to the non-deterministic nature of these technologies. The objective is to develop reference guides and standards for the integration, evaluation, and certification of AI in aeronautical systems.

Targets and scope

Embedded systems (aircraft, drones) and ground-based systems (ATM/ANS), with European and international reach.

Description

Working group led by Airbus, Thales, and Collins Aerospace. Work organized into sub-groups (applications, ML data management, verification, security, human factors).

Results

- The Statement of Concerns (ER-022, 2021) formally identified the main concerns related to the use of AI in aviation.
- The Taxonomy in AI (ER-027, 2024) defines a structured framework for categorising different types of AI.
- The Use Cases Considerations (expected in 2026) guides the evaluation of concrete applications of AI in the aeronautical sector.
- The Process Standard for Development and Certification/Approval of aeronautical Products implementing AI (ED-324, expected in 2026).

Conclusion and outlook

WG-114 is a key initiative for building a robust normative framework around aeronautical AI. The publication of ED-324 will be decisive in consolidating the trust and regulatory acceptability of AI in Europe.

5.1.3 Thales publication references

1. Joint report Thales & CESIN – L'IA & les RSSI

Date: March 2025

Participants: Thales, CESIN

Description: White paper presenting a study conducted among CESIN member CISOs, exploring the impact of AI on their profession and the associated governance challenges. It sets out challenges and recommendations for integrating AI throughout the project lifecycle.

2. TAID White Paper – Trustworthiness for AI in Defence

Date: May 2025

Participants: Center of Excellence for Cyber, EDA, AMIAD, Thales, Capgemini, European industrial players, academic and institutional experts

Description: Document produced by the TAID working group, aimed at analyzing and recommending the key aspects of responsible, ethical, and reliable AI system development for European defence.

3. DEEL Project – Dependable & Explainable Learning

Date: March 2021

Participants: IRT Saint-Exupéry, ANITI (Toulouse), IVADO et CRIAQ (Montreal), industrial players (aeronautics, space, automotive) such as Thales, Airbus…

Description: Consortium of 60 experts (researchers, data scientists, security and certification engineers) based across Toulouse and Quebec.

Objective: Development of methods guaranteeing the reliability and explainability of AI for industrial critical systems.

4. Collective Work – Trustworthy AI in defence

Date: June 2025

Participants: Center of Excellence for Cyber, AMIAD, Thales, Capgemini, European industrial players, academic and institutional experts

Description: Collective publication analyzing the strategic, technical, ethical, and legal challenges of Trustworthy AI, in service of digital and military sovereignty.

5. Position Paper – AI for critical systems

Date: November 2024

Participants: Thales

Description: Document presenting the vision of AI for critical systems as a controlled support for human decision-making, integrating stringent requirements for security, robustness, governance, and regulatory compliance.

6. Position Paper – End-to-End Trustworthy AI Engineering Lifecycle for critical systems

Date: January 2026

Participants: Thales

Description: Document that presents a vision of AI in critical systems, in which reliability, security, and transparency are indispensable. It shows how Thales designs its solutions end-to-end, from data to algorithms, through to validation and operational use, while following the industrial chain of trust.

6. Conclusion and outlook

Trustworthy AI is not limited to a compliance requirement; it is an **invitation to rethink our approaches to innovation and collaboration.** It represents a **proactive and systemic approach** integrating ethical, social, and environmental principles at the very heart of its development and use.

The challenge is not to define a model alone, but to contribute to a collective, evolving, and inclusive effort, in order to build **ethical and responsible AI.** This translates into adapting our governance models, strengthening dialogue between public, private, and academic actors, and by defining **the responsibility of stakeholders** throughout the lifecycle of AI systems. Clarifying the responsibility, accountability, and liability of actors during the development and deployment of AI systems lies at the heart of Trustworthy AI, so that stakeholders and users understand their rights and obligations.

The implementation of new regulations and the acceleration of AI use call for **continuous vigilance and reflection in order to pursue the construction of Trustworthy AI**, so as to meet the challenges associated with technological and societal developments. The emergence of AI agents, capable of executing complex tasks autonomously and of interacting with other systems, represents an example of these new challenges.

These systems indeed require enhanced supervision in order to ensure that their actions remain aligned **with human intent and societal values.** This context further highlights **the question of responsibility:** that of designers in defining the need, the limits, and the possible uses, and that of users in their engagement with and control over the conditions of use.

7. About Thales

Today, Thales brings together over 800 AI experts within its internal dedicated organization, cortAIx, created in 2024. These experts are divided into three main teams:

- **Labs:** focused on fundamental research and model development.
- **Factory:** responsible for tool chains, cybersecurity, and critical infrastructure.
- **Sensors:** integrating AI into embedded and edge systems.

Together, these teams ensure that AI capabilities are not only state-of-the-art but also operationally viable, secure, and compliant with sovereignty requirements. Thales' cortAIx teams are strategically located across France, the United Kingdom, Canada, Singapore, Germany and the United Arab Emirates, reflecting the company's global footprint and its commitment to supporting regional sovereignty through local expertise.

cortAIx
Artificial Intelligence by THALES

8. Resources

Regulatory / Standards level

- Council of Europe - Framework Convention on Artificial Intelligence — Link
- European AI Act, *2024* — Link
- RGPD, *2016* — Link
- European Cyber Resilience Act, 2024 — Link
- CEN-CENELEC JTC 21 - EU AI Act Standards, *2025* — Link
- ISO/IEC 42001 - AI Management system, *2023* — Link
- ISO/IEC 38507 - Governance implications of the use of artificial intelligence by organizations,*2022* — Link
- ISO/IEC TR 24028 - Overview of trustworthiness in artificial intelligence, *2020* — Link
- ISO/IEC 42005 - AI system impact assessment, *2025* — Link
- ISO/IEC CD 42105 - Guidance for human oversight of AI systems (under development) — Link
- ETSI - Technical Committee (TC) Securing Artificial Intelligence (SAI) TR 104, *2024-2025* — Link
- ISO/IEC 27001 - Information security management systems - Requirements, *2022* — Link
- ISO/IEC 23894 – AI Guidance on risk management, *2023* — Link
- Loi AGEC, *2020* — Link
- Loi REEN, *2021* — Link
- Loi Climat et Résilience, *2021* — Link

Recommandation level

- OCDE - Principles for Trustworthy AI, *2024* — Link
- Commission Européenne – Assessment List for Trustworthy Artificial Intelligence (ALTAI), *2021* — Link
- IEEE - Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, *2019* — Link
- HLEG UE - Ethics Guidelines for Trustworthy AI, *2019* — Link
- CNIL - Ensuring the security of AI system development, *2025* — Link
- ANSSI - Recommandations de sécurité pour un système d'IA générative, *2024* — Link
- ENISA - Securing Machine Learning Algorithms, *2021* — Link
- NIST - AI Risk Management Framework, *2023* — Link
- CNIL - AI and GDPR, *2025* — Link
- CNIL – AI how-to sheets, *2024* — Link
- UNESCO - Ethics of Artificial Intelligence, *2021* — Link
- NIST - Four Principles of Explainable Artificial Intelligence, *2021* — Link
- CNPEN – Opinion n°7: Generative artificial intelligence systems: ethical issues, *2023* — Link
- INR - RIA31 Guide IA Ethique et Responsable, *2024* — Link
- Arcep / Arcom / ADEME - Référentiel général d'écoconception de services numériques (RGESN), *2024* — Link
- AFNOR / Ecolab - Spec 2314 - IA Frugale, *2024* — Link
- CEN-CENELEC - Standards in support of trustworthy environmental claims, *2025* — Link
- Paris AI Action Summit - Standardization for AI Environmental Sustainability, *2025* — Link

Professional level

- Hub France IA - Livre Blanc IA éthique en pratique (2e version), *2024* — Link
- Hub France IA / Wavestone – Notice Premier pas vers l'IA de confiance, 2025 — Link
- Impact AI - Les briefs de l'IA Responsable, *2024* — Link
- CIGREF / Numeum - Guide de mise en œuvre de l'AI Act, *2025* — Link
- Confiance.ai - Livre Blanc « Towards the engineering of trustworthy AI applications for critical systems », *2022 et 2024* — Link
- ITRE (European Parliament) - Interplay between the AI Act and the EU digital legislative framework, *2025* — Link
- Lefebvre Dalloz - AI ACT, le nouveau cadre juridique européen de l'IA, *2025* — Link
- Microsoft - Whitepaper « Accelerate AI transformation with strong security », *2022* — Link
- Microsoft - Secure AI: processes for securing AI, *2025* — Link
- Microsoft – Responsible AI Impact Assessment Template, *2022* — Link
- Google – Secure AI Framework, 2023 — Link
- The Digital Economist – The ROI of AI Ethics, *2025* — Link
- Numeum - Ethical AI (livret pédagogique), *2024* — Link
- Numeum / G9+ / CIGREF / Planet Tech Care / Hub France IA – AI for Green & Green AI, *2024* — Link
- Green IT - Environmental and health impacts of AI worldwide, *2025* — Link
- Green IT - Référentiel de maturité/ 74 bonnes pratiques pour un numérique plus responsable, *2022* — Link
- The Shift Project - AI, data, and computing: shaping infrastructures for a decarbonised world, *2025* — Link

9. Acknowledgements

The completion of this article, which is the result of an in-depth exploration of the challenges and prospects of Artificial Intelligence, would not have been possible without the contribution and commitment of the Thales expert community.

We wish to express our gratitude to the professionals, researchers, and experts who generously shared their knowledge and insights, making a significant contribution to the development of this work. Their expertise has enabled us to provide a versatile and nuanced perspective on a subject that is constantly evolving.

We would like to extend our special thanks to our Thales experts:

- For their expertise in AI and their cross-functional contributions:

Thomas **DELAVALLADE**
Benoit **HUYOT**
Nicolas **MUSEUX**

- For analyses of the security, safety and cyber aspects of AI:

Olivier **BETTAN**
Joseph **MACHROUH**
Jean-Marie **KEBALO**

- For insights into the aspects of robustness, transparency, explainability and trust in AI:

Patricia **BESSON**
Florence **DE GRANCEY**
Nick **ERNEST**

- For contributions on the ethical aspects of AI:

Benoit **HUYOT** - Design Authority for AI
Pierre-Etienne **LEGOUX** - Design Authority for AI

- For insights into the sustainability aspects of AI:

Remi **BARRERE**
Thomas **HEUSSE**
Aurore **TUAL**
François **MORIAMEZ**
Marc **HEUDE**
Pierre **GIRARD**

- For further information on the legal aspects:

Victor **CART**

Their precision and rigor have been essential in ensuring the integrity and relevance of the information presented.

We also wish to thank our Data & AI consulting practice leader Jean-Paul **LEROUX** and also, for their support for this project, to:

Hélène **BRINGER**

Juliette **MATTIOLI**

Juliette **ROUILLOUX-SICRE**

And David **SADEK**

THALES

Building a future we can all trust

4, rue de la Verrerie 92190
Meudon FRANCE

Tél. + 33(0)1 57 77 80 00

www.thalesgroup.com