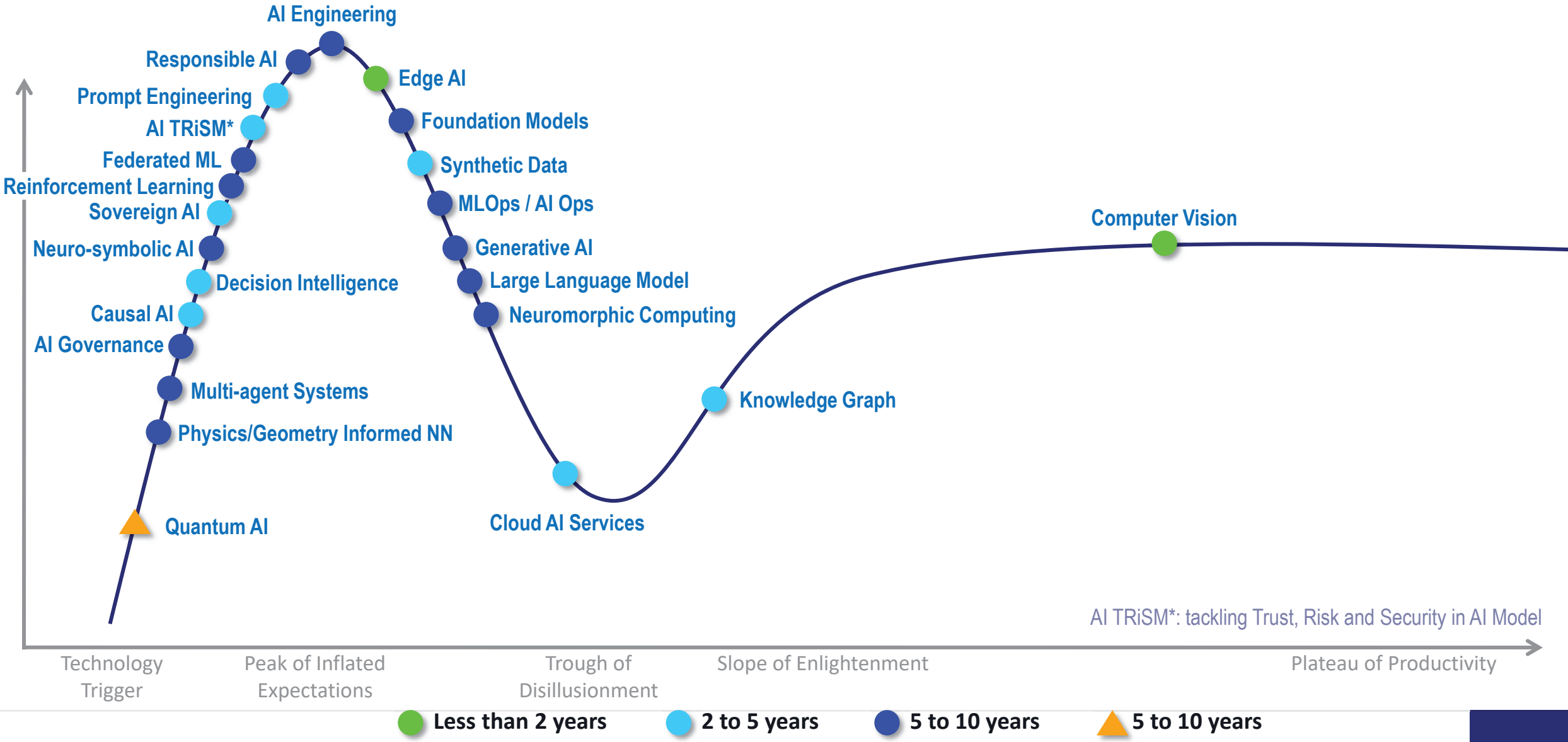
THALES
Building a future we can all trust

AI future main challenges

Juliette MATTIOLI

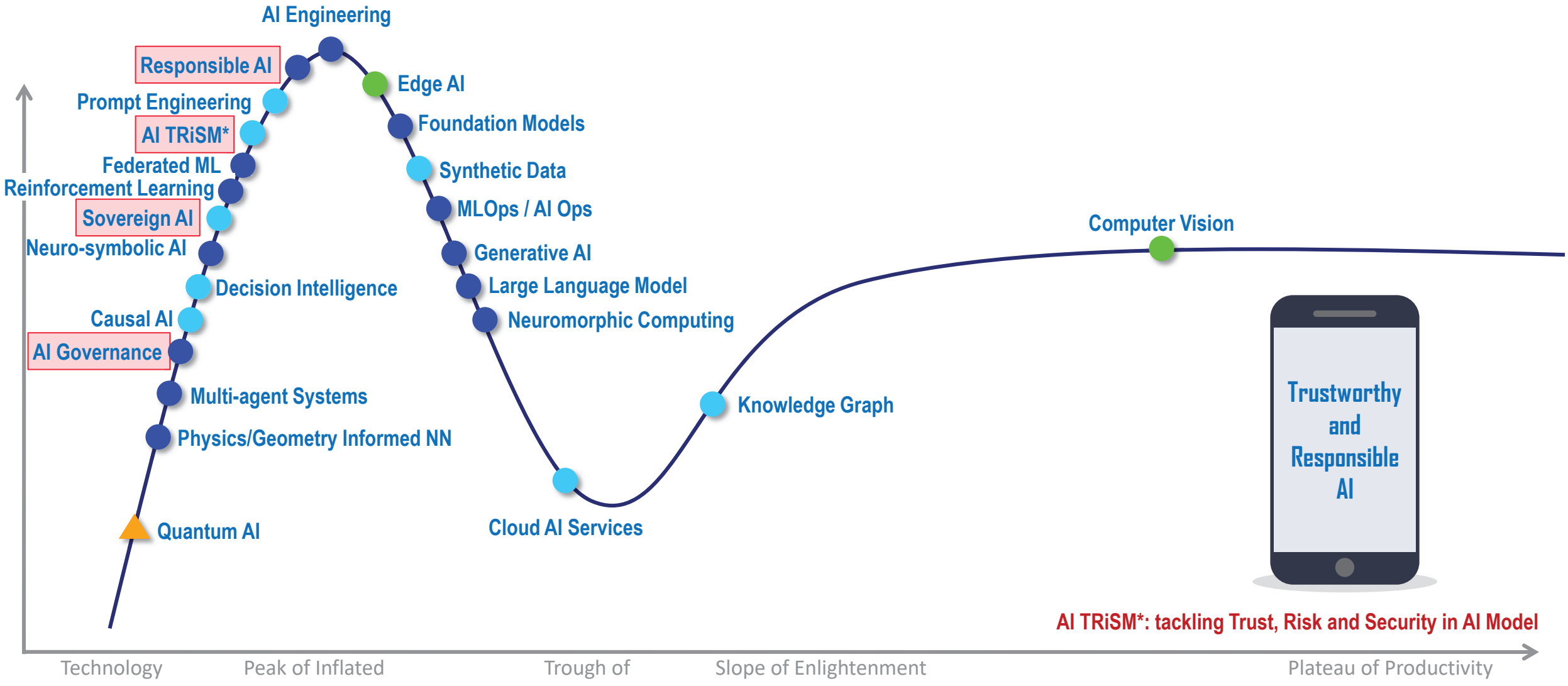
www.thalesgroup.com

AI Main Technology Trends (mainly from 2024 AI Gartner hype curve)



{OPEN}

AI Main Technology Trends: Trustworthy and Responsible AI



{OPEN}

Trustworthy and Responsible AI

Validity

To guaranty that an AI-based system will do what it is meant to do, **all** what it is meant to do and **only** what is meant to do

Security

To ensure **robustness and resilience** to adversarial conditions, such as decoying and cyber-attacks

Explainability

To be able to provide **human-level, understandable and context-relevant** justifications and explanations

Responsibility

To be compliant with **ethical, legal** and **regulatory** frameworks

Thales TrUE AI – Transparent, Understandable, Ethical

{OPEN}

Data & AI Regulation

GDPR *Applicable since May 2018*	EU DATA GOVERNANCE ACT *Applicable September 2023*	DATA ACT *Applicable September 2025*	AI ACT *Applicable August 2026 for "High risk" systems*	AI liability package *Still under discussion*
Accountability Right to access Breach notification Harmonised fines	Labelled data intermediaries EU Governance on data Data Sovereignty	Access to data generated by IoT to users Data sharing with public sector for emergencies Easier change of cloud provider Data Sovereignty	Risk based approach Robustness, accuracy and cybersecurity Transparency and information Governance and quality of datasets Human oversight Conformity assessment	Creation of an extra-contractual liability regime to repair prejudice caused by AI Revision of the product liability directive to include AI within it scope
Ensure that companies handle personal data in a responsible and accountable way, and that individuals have greater control over their data	Regulation that aims at facilitating the data sharing across EU	Regulation that aims at facilitating the data sharing across EU	Requirements and obligations tailored on a 'risk-based approach' and the role in the AI value chain	Make sure that persons harmed by AI enjoy the same level of protection as persons harmed by other technologies

{OPEN}

Within the standard landscape

ISO/IEC 2382:2015
Information technology — Vocabulary

ISO/IEC JTC 1/SC 7
Software & systems engineering

ISO/IEC 25030:2019
Systems and software quality requirements and evaluation (SQuaRE)

ISO/IEC/IEEE 15288:2023
SYS & SW Engineering — System life cycle processes

ISO 31000:2018
Risk management — Guidelines

ISO/IEC JTC 1/SC 39
Sustainability, IT & data centers

NATO (STANAG)

Trustworthy & Responsible AI

IEEE 7000s

OMG Standards Development Organization

****

ISO/IEC 22989:2022
AI concepts & terminology

ISO/IEC 42001:2023
AI Management system

ISO/IEC 23053:2022
Framework for AI Systems using ML

ISO/IEC 5392:2024
AI reference architecture of knowledge engineering

ISO/IEC TR 5469:2024
Functional safety & AI systems

AI for …
AI for Space
AI for Aeronautics
EUROCAE WG114
ARP6983

ISO/IEC JTC 1/SC 42
Artificial intelligence

ETSI

ISO/IEC 5338:2023
AI system life cycle processes

ISO/IEC 5339:2024
Guidance for AI applications

ISO/IEC TR 24028:2020
Overview of trustworthiness in AI

ISO/IEC AWI TS 5471
Quality evaluation guidelines for AI systems

ISO/IEC 23894:2023
Guidance on AI risk mgt

Key Principles of Trustworthy and Responsible AI

{OPEN}

Trustworthy & Responsible AI (1/3)

> Transparency & Explainability

- AI systems should be **transparent**. AI providers should provide **clear information about the system's capabilities and limitations**, as well as the **data sources used to train it**.

> Privacy & data protection

- AI system providers and developers should be designing AI systems with **data privacy** and **data protection** in mind. The datasets used to train AI systems should be **properly governed.**

> Robustness and safety

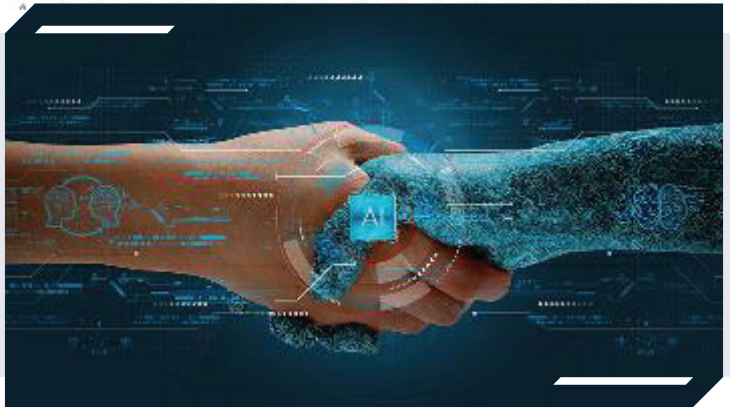
- Providers and developers of AI systems should be designing AI systems to **work well**, be **predictable**, and **be safe to use**. AI providers should ensure that their systems comply with **quality management systems**.

{OPEN}

Trustworthy & Responsible AI (2/3)

> Accountability

- Organizations are required to establish clear lines of accountability and oversight for AI development, deployment, and use. This involves defining **roles and responsibilities**, establishing **governance frameworks**, and enforcing mechanisms for **auditing, monitoring, and handling the performance and impact** of AI systems.



> Collaboration and Multi-Stakeholder Engagement

- By gathering diverse perspectives, organizations can make more **informed decisions**, address **societal concerns**, and ensure that AI benefits a wide range of stakeholders.

> Human-Centric Design

- AI systems should **assist humans** in decision-making, and humans should be able to override decisions made by the system.

{OPEN}

Trustworthy & Responsible AI (3/3)

> Diversity, non-discrimination, fairness

- Developers and providers of AI systems should be designing AI systems to **avoid discrimination and bias** and **promote diversity**. Providers should carefully examine data sources for bias and use proper measures to **mitigate any potential biases**.

> Social and environmental impacts

- Designers of AI systems should be designing AI systems to contribute to **sustainable and inclusive growth, social progress, and environmental well-being**. Providers should consider the potential impact of AI systems on society and the environment

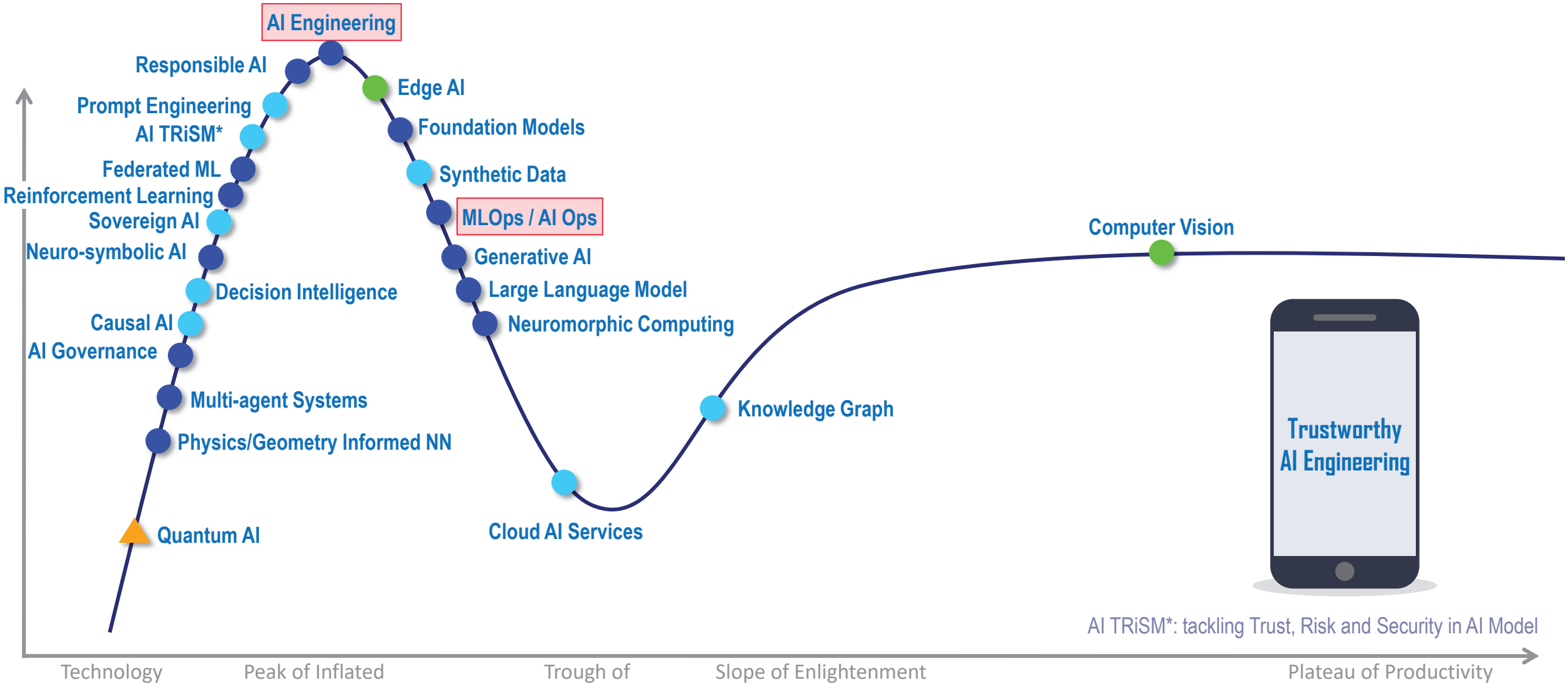
> AI TRiSM: AI trust, risk and security management

- To ensure **AI governance**, trustworthiness, fairness, reliability, robustness, efficacy and data protection.
- Compliance in a global landscape (regulation, standards….)

> Sovereign AI

{OPEN}

AI Main Technology Trends: Trustworthy AI Engineering



Less than 2 years | 2 to 5 years | 5 to 10 years | 5 to 10 years

{OPEN}

AI/ML deployment induces some (engineering) challenges…

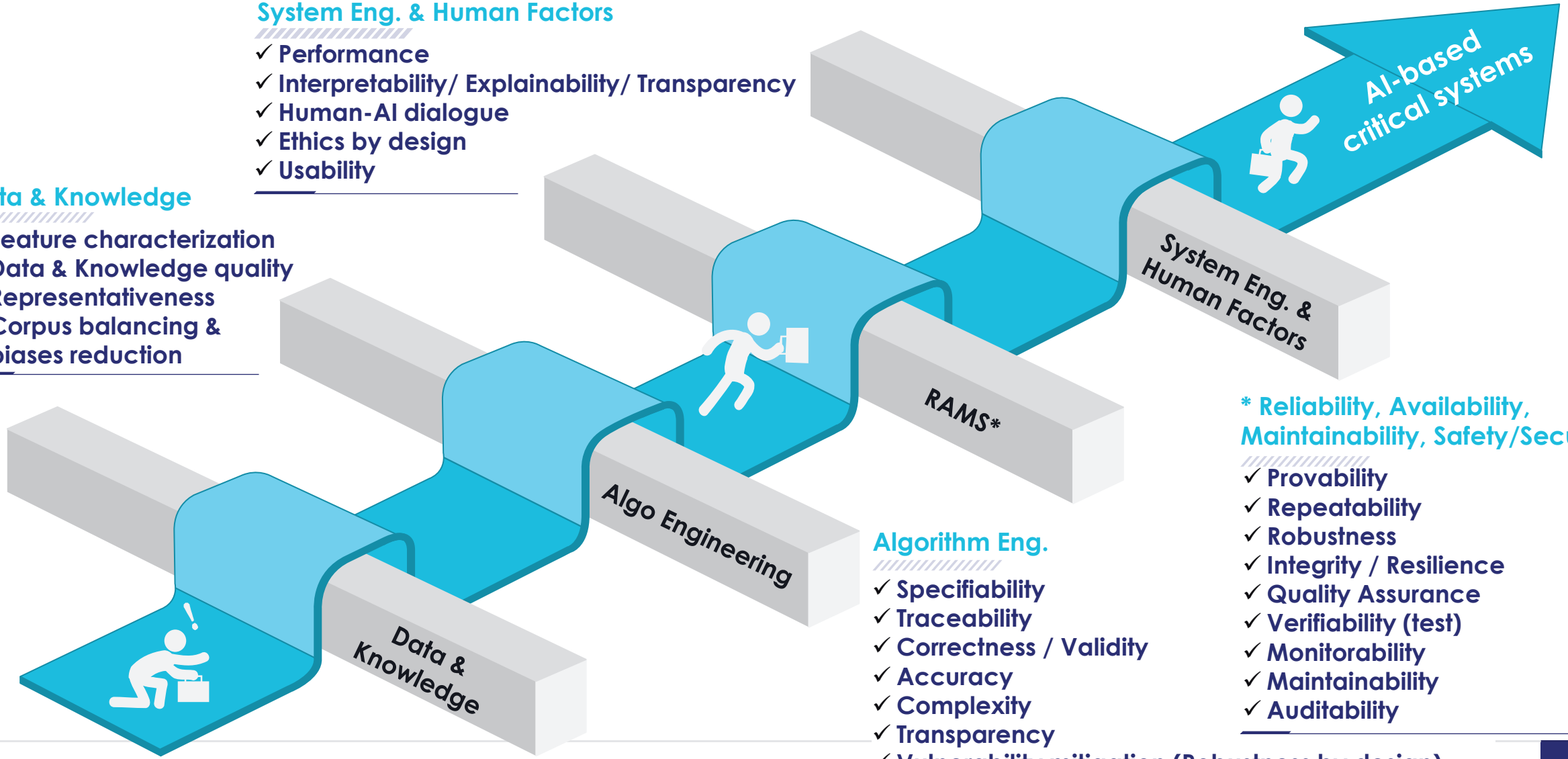
(confiance.ai & confiance IA)

System Eng. & Human Factors

- ✓ Performance
- ✓ Interpretability/ Explainability/ Transparency
- ✓ Human-AI dialogue
- ✓ Ethics by design
- ✓ Usability

Data & Knowledge

- ✓ Feature characterization
- ✓ Data & Knowledge quality
- ✓ Representativeness
- ✓ Corpus balancing & biases reduction



Algorithm Eng.

- ✓ Specifiability
- ✓ Traceability
- ✓ Correctness / Validity
- ✓ Accuracy
- ✓ Complexity
- ✓ Transparency
- ✓ Vulnerability mitigation (Robustness by design)

* Reliability, Availability, Maintainability, Safety/Security

- ✓ Provability
- ✓ Repeatability
- ✓ Robustness
- ✓ Integrity / Resilience
- ✓ Quality Assurance
- ✓ Verifiability (test)
- ✓ Monitorability
- ✓ Maintainability
- ✓ Auditability

Trustworthiness in AI-based critical system impacts the overall engineering lifecycle

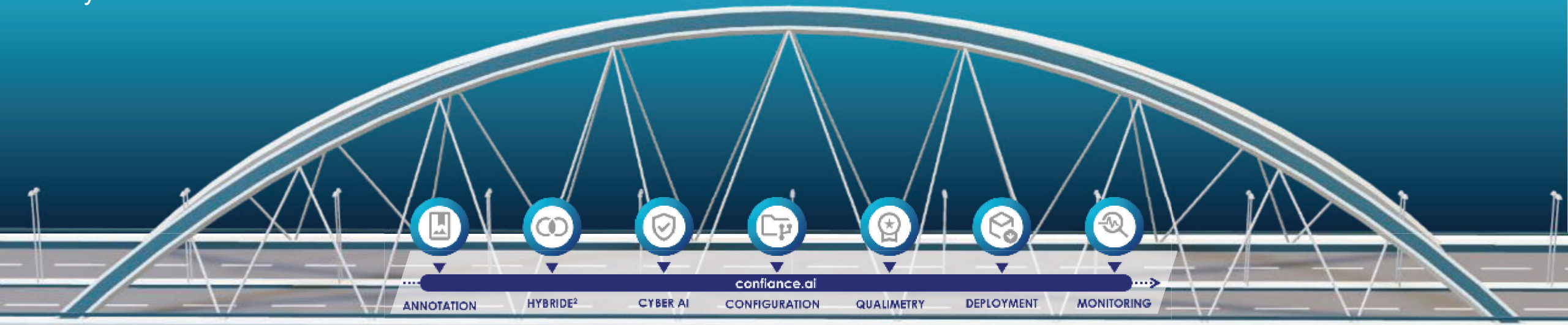
From requirements & specifications

→ Stakeholder reqs.
→ Sys. & AI/ML specs.
→ Sys & AI/ML archis

MLOps tool-chain

To system deployment & maintenance

→ Monitoring
→ Toward qualification and certification



AI engineering: a necessarily condition to deploy trustworthy AI

Operational Design Domain (ODD)

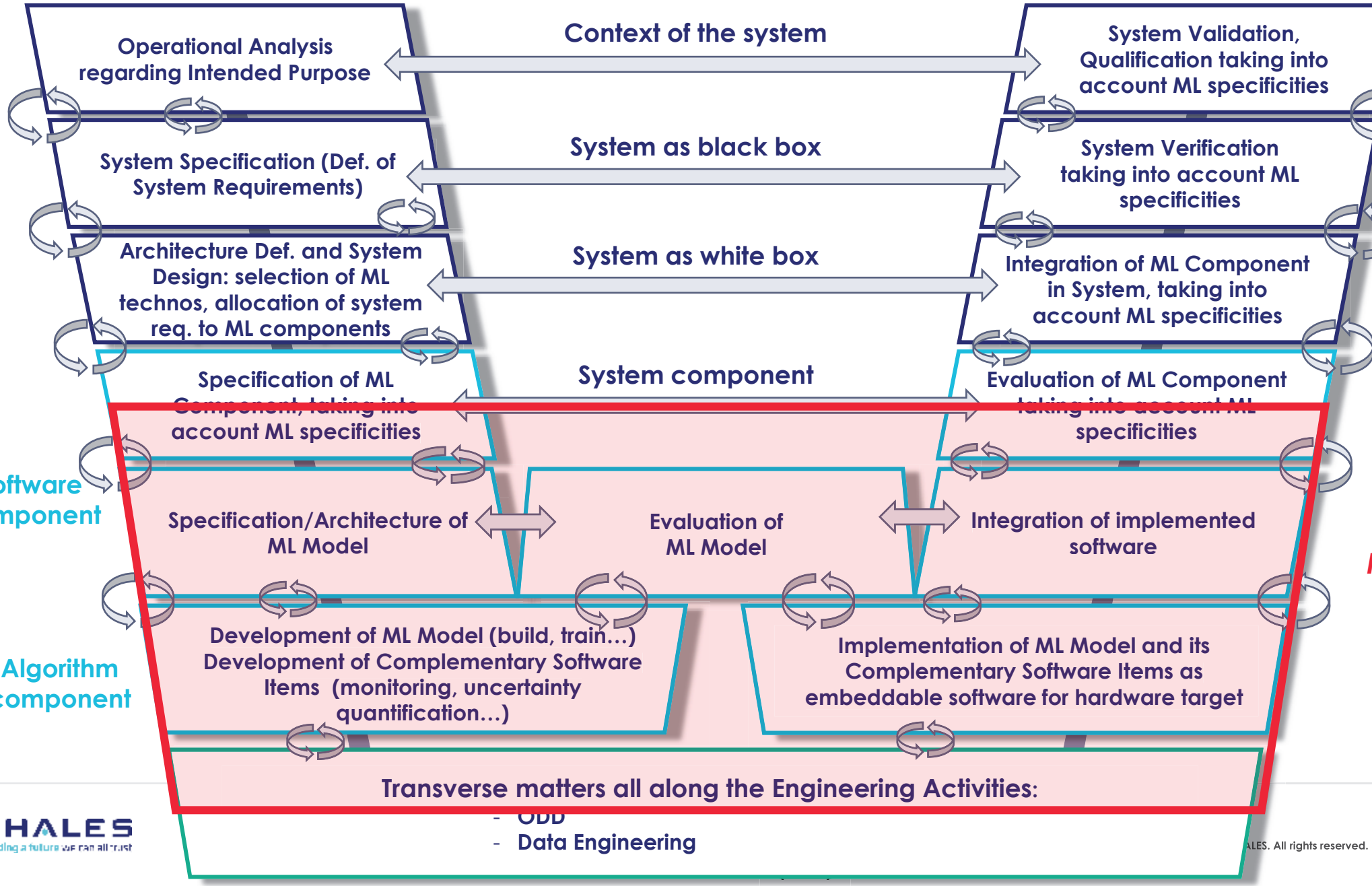
Validation

Verification

Operational Analysis regarding Intended Purpose

To ensure qualification and compliance with regulations (e.g. AI Act) **and standards** (e.g. ARP6983 in aeronautics)

Systems/Software/Algorithm/Data Engineering lifecycle to design a ML-based System



Leveraging the Concepts of Design Assurance for Neural Networks – CoDANN [EASA]

MLOps Chain

MLOps to support Data-driven AI deployment

> ML variety (according to the problem)

- To consider several types of ML (not only supervised ML)

> Operationalization

- To mitigate the technical debt of involved engineering disciplines (algo, systems, safety, security and human factors), their inherent complexity, to the emergent risks induced by ML deployment within critical systems Governance

> Managing Model Drift

- MLOps engineers must ensure that models are monitored and retrained regularly to maintain their accuracy and mitigate the degradation of model performance over time

> Trustworthiness

- To accelerate the journey of trustable ML deployments into critical systems

> Scaling Machine Learning Infrastructure

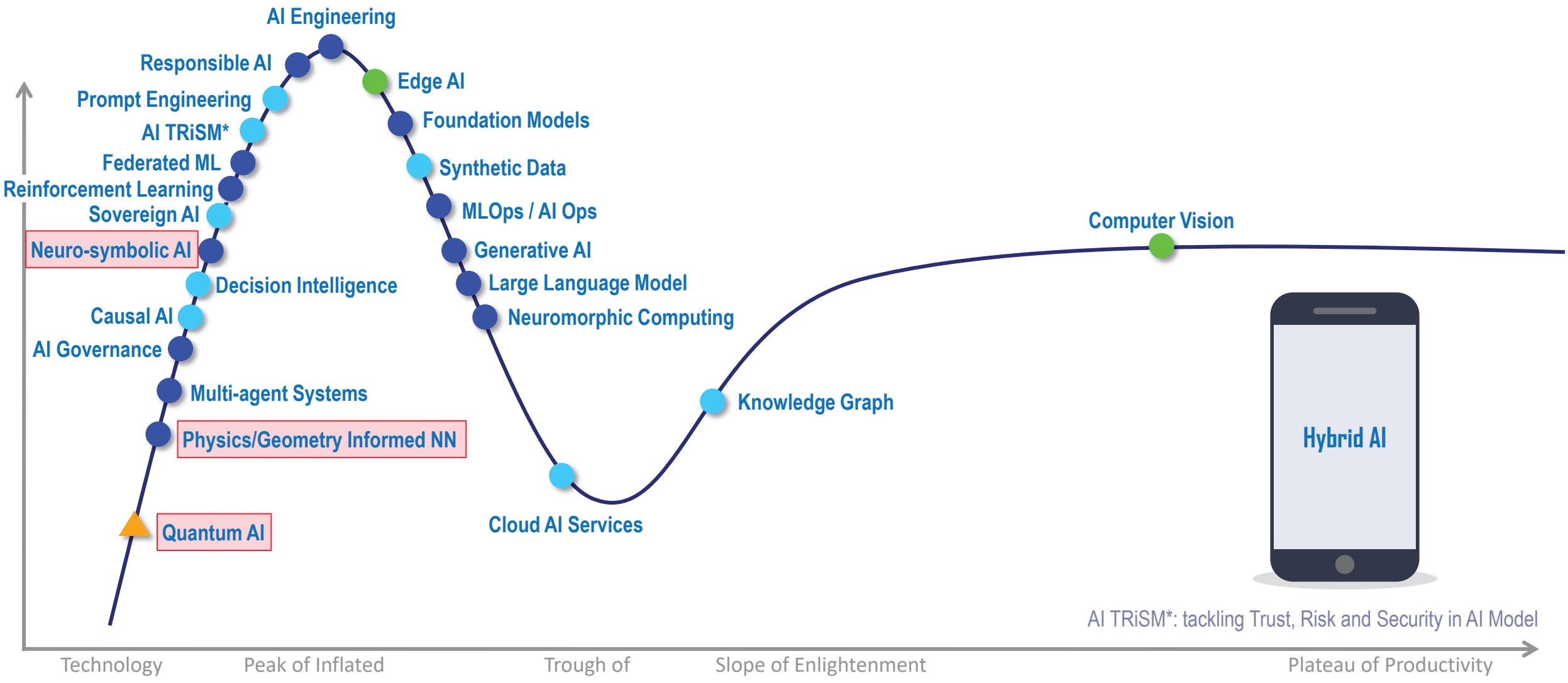
- To ensure that models can handle increased loads, integrating with multiple data sources, and deploying models across distributed environments.

> Experiment Tracking and Reproducibility

- Experiment tracking systems must be in place to ensure that different iterations of models can be compared and retrained as needed, to guaranty that ML experiments are well-documented and reproducible.

{OPEN}

AI Main Technology Trends: Hybrid AI



{OPEN}

Hybrid AI

Rationale

- Capture the strengths of data-driven and knowledge-based AI.
- Integrate seamlessly data and mathematical physics models, in uncertain and high-dimensional contexts.
- Support synergy, symbiosis, and augmentation of human and AI.
- More explainable; Easier to validate; Frugal: requires less data; Faster to run.

Approaches

- **Combination of methods and techniques from AI subfields,** possibly enhanced by knowledge (Math, Physics, Geometry…)
- **Neuro-symbolic ML**
- **Knowledge Informed NN** Physics: PINN; Geometry: GINN
- **ML & Reasoning by analogy**
- **Quantum AI**

Technical Issues

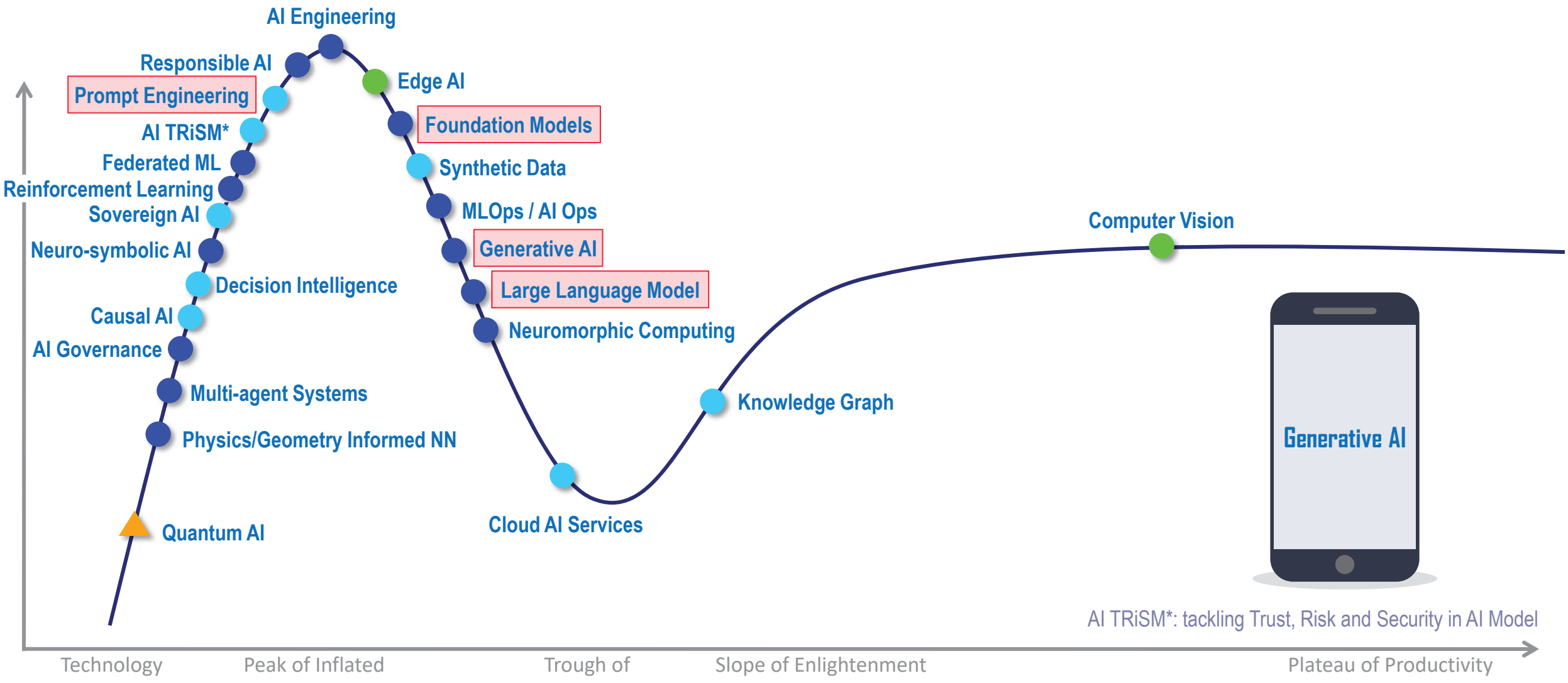
- Development of AI algorithms which are;
- Robust to limited data
- Adaptable to environment with size, weight and communication constraints.

Use-cases

- Descriptive and prescriptive analysis for maintenance operations
- Fisheye Image Processing
- Forecast a cyber attack plan
- Radar Doppler Signal processing
- Design and monitoring of physical systems
- Landing recommendation (approach, landing, taxi, gate)

{OPEN}

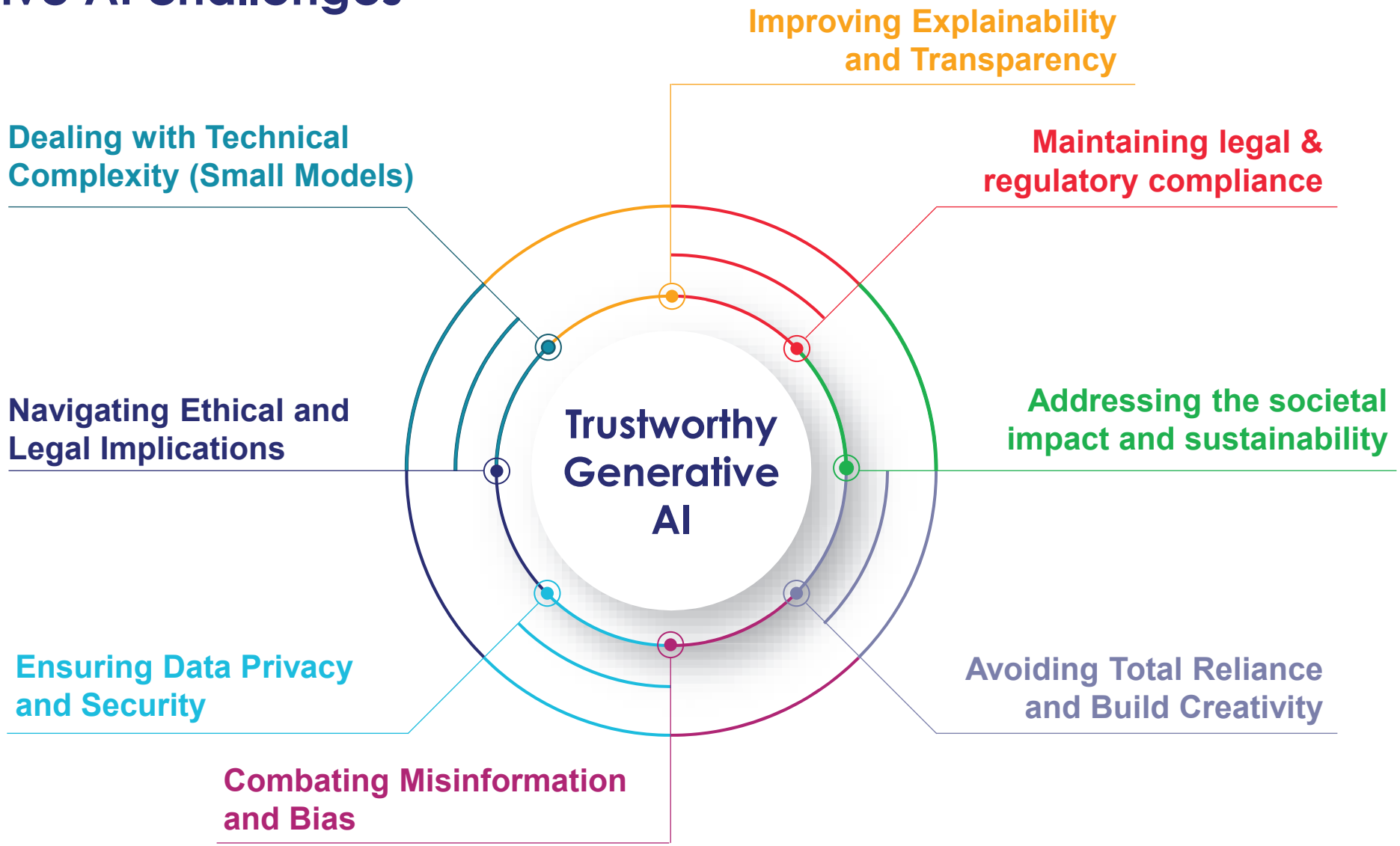
AI Main Technology Trends: Generative AI



Less than 2 years | 2 to 5 years | 5 to 10 years | 5 to 10 years

{OPEN}

Generative AI challenges



Wrap-up: Toward Responsible/Certifiable AI

Validity

To guaranty that an AI-based system will do what it is meant to do, all what it is meant to do and only what is meant to do

Transparency & Explainability

To be able to provide human-level, understandable and context-relevant justifications and explanations

Accountability & Responsibility

To be compliant with ethical (including sustainability), legal and regulatory frameworks

Security & Robustness

To ensure robustness and resilience to adversarial conditions, such as decoying and cyber-attacks

Data & Privacy Governance

(incl. Data Quality)

Dependability & Safety

To control risk of unacceptable failures and functional insufficiencies

{OPEN}

THALES
Building a future we can all trust

Thank you

www.thalesgroup.com