THALES
Building a future we can all trust

Quels enjeux pour une IA de confiance ?

Juliette MATTIOLI
Thales AI Fellow

www.thalesgroup.com

l'IA est « partout »

> Internet et assistant personnel

- Anti-spam, correcteur orthographique, assistants vocaux

> Santé

- Aide au diagnostic (Mélanome cancéreux, cancer des poumons; …), Aide à la conception de molécules médicamenteuses

> Commerce et marketing

- Systèmes de recommandation (Netflix, Amazon, Ebay), Conseiller en vente (Sephora, H&M)…

> Finance, banques & assurances

- Octroi de crédit, Support client, Conseil et Vente,

> Transport

- Google Maps, Waze, véhicules autonomes, applications de covoiturage

> Environnement

- Régulation de la consommation énergétique de bâtiments intelligents, Observation de la terre, Optimisation de la gestion de l'eau pour l'irrigation

> Défense, Sécurité

- Soldats augmentés, Vidéo-surveillance, Identification par reconnaissance biométrique

REF Thales/PCC - 2025/JM,30 – 13 juin 2025

AI applications across Thales markets

AEROSPACE

Eco-Friendly Operations

Air traffic Mgt

Digital Pilot Notebook

Disruptive Pilot Training & Assessment

DEFENSE

Automatic (any milieu) Target Detection & Recognition

Digital Crew

Electronic Warfare

Mission Risk Analysis

Tactical Decision Optimization

Indoor or GNSS denied geolocation

Collaborative Combat

DIGITAL IDENTITY & SECURITY

Anomaly Detection

Liveness Detection System & Biometry

Fraud Mgt

AI for Cybersecurity

Cybersecurity for Surveillance

SPACE

Imagery for Earth Observation

Satellite tele measures analysis

Satellite / Constellation Mission planning

Automatic guidance, debris avoidance

THALES GROUP

AI for Engineering

Trust AI Engineering

Smart In Service Support incl. HUMS

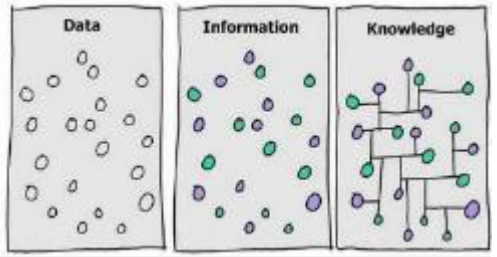
Knowledge Mgt

Business/Techno Intel.

Yield mgt & Industry 4.0

IA : Capacités cognitives dans un système artificiel

Inputs



Donnée

Information

Connaissance

AI Algorithm

Outputs

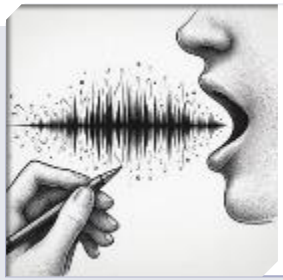
- **Perception**
  - informations riches, complexes et imparfaites
- **Apprentissage**
  - à partir d'exemples
- **Abstraction**
  - création de sens
- **Raisonnement**
  - Découverte de connaissances, planification et décision
- **Communication**
  - dialogue naturel
- **Action**
  - pour atteindre un objectif rationnel

REF Thales/PCC - 2025/JM,30 – 13 juin 2025

IA dirigée par les données

Quelques exemples d'applications

Segmentation d'images / Détection et reconnaissance d'objets



Reconnaissance de la parole

Prédiction du trafic

Prédiction météo

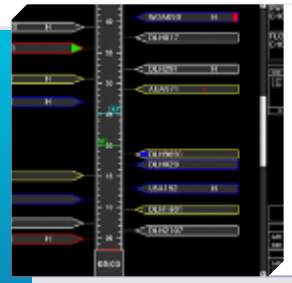
Détection d'anomalies et de comportements anormaux

Reconnaissance biométrique pour le contrôle aux frontières

REF Thales/PCC - 2025/JM,30 – 13 juin 2025

IA à base de connaissance

Quelques exemples d'applications



Planification de la gestion du trafic aérien

Gestion des ressources et de la supply-chain

Maintenance préventive et prescriptive

Gestion de la qualité et des rendements pour la production

Optimisation logistique

IA Générative

Quelques exemples d'applications

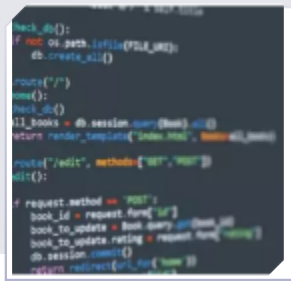
Génération de texte et image

Traduction automatique

Génération synthétique de données

Assistant pour le maintien en condition opérationnelle

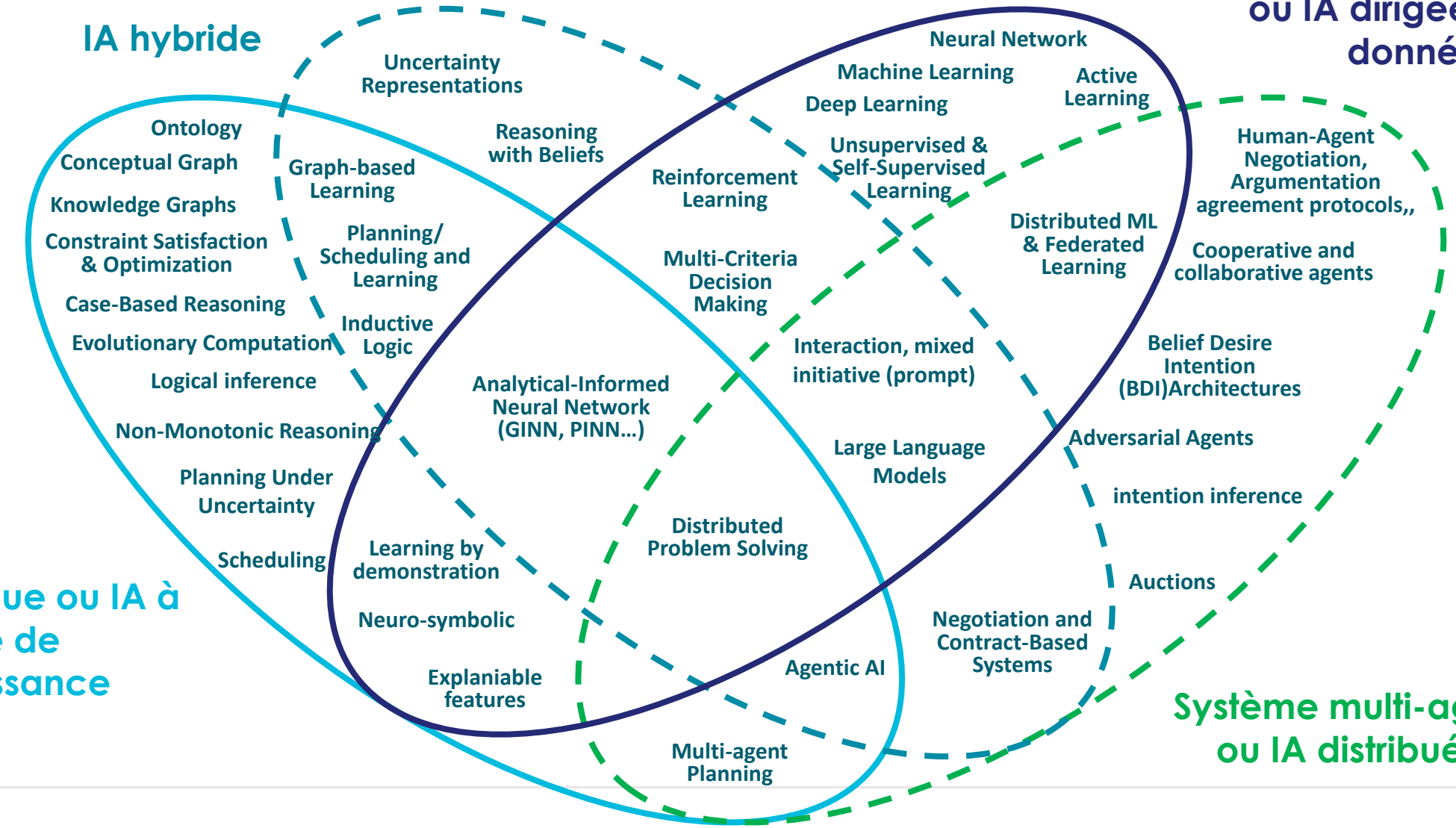
Génération automatique de rapport



Production automatique de code

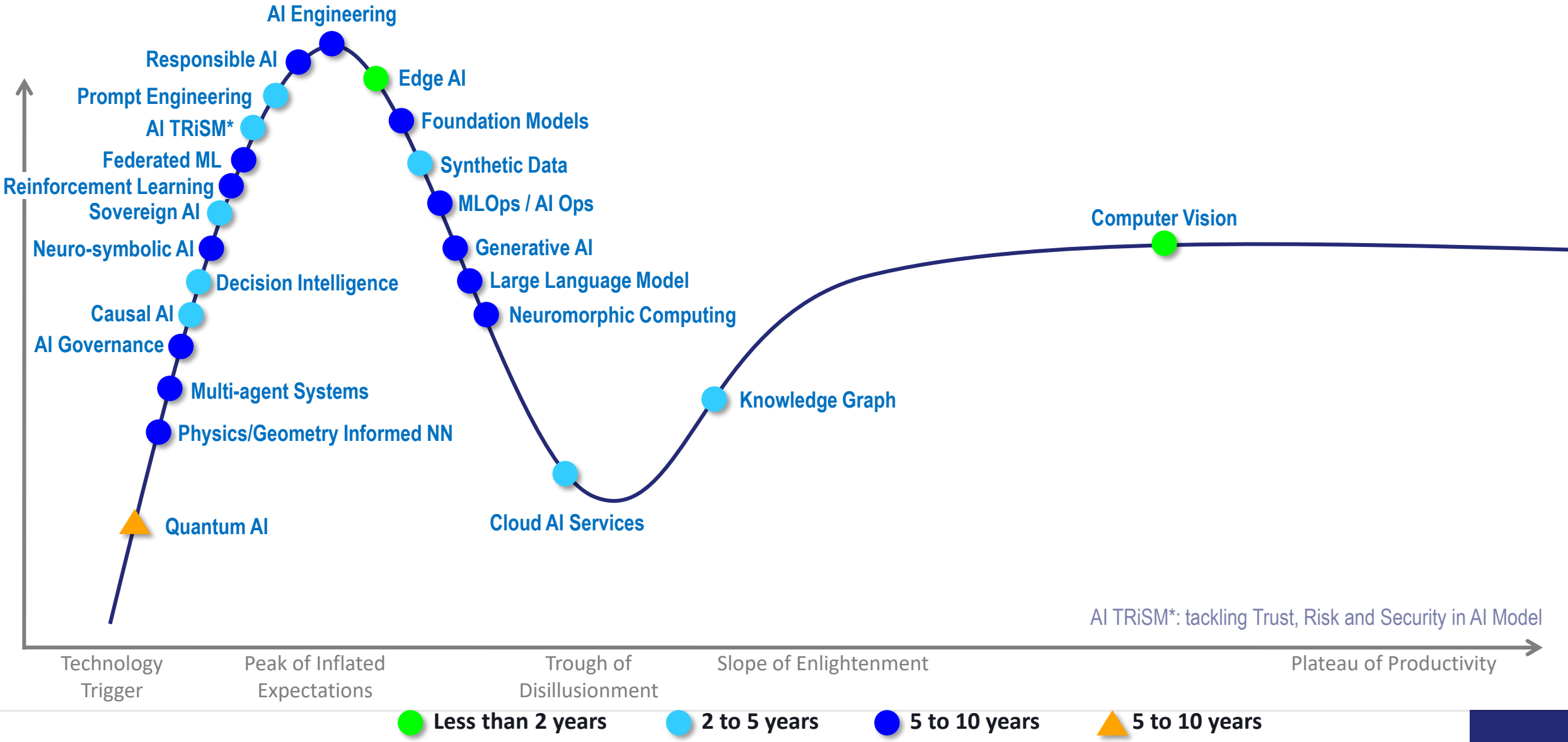
REF Thales/PCC - 2025/JM,30 – 13 juin 2025

Une zoologie d'algorithmes d'IA



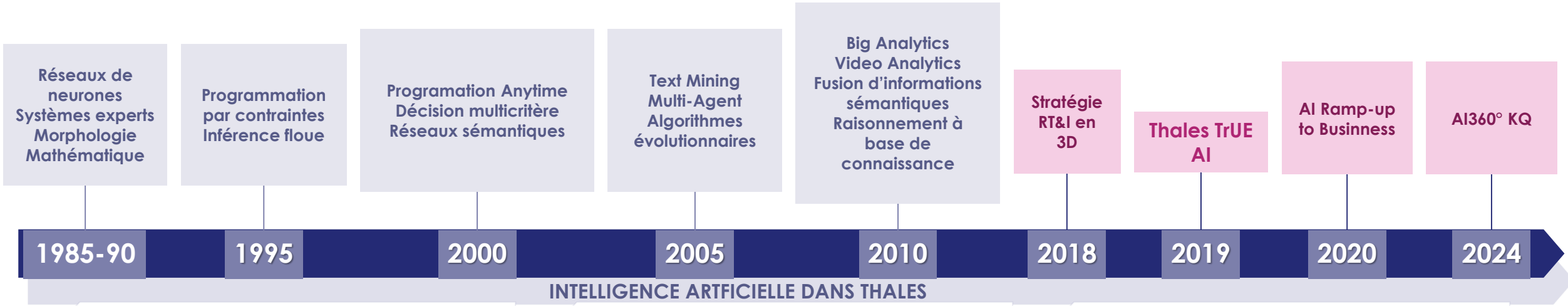
REF Thales/PCC - 2025/JM,30 – 13 juin 2025

Les grandes tendances en IA (2024 AI Gartner hype curve)



REF Thales/PCC - 2025/JM,30 – 13 juin 2025

IA @ Thales



Thales TrUE AI

8 technologies différentiantes

Communauté AIDA

Collège AIDA & Thales School of AI

~100+ PhD

800+ IA ingénieurs & scientifiques

250+ publications scientifiques (depuis 2020)

+250 brevets en IA (depuis 2018)

Charte éthique du numérique de Thales (Oct 2022)

L'ingénierie de l'IA de confiance

Le déploiement de l'AI Act

IA de Confiance : « *Thales TrUE AI – Transparent, Understandable, Ethical* »

Validité

Garantir qu'un système d'IA fait ce qu'il doit faire, **tout ce qu'il doit faire et seulement ce qu'il doit faire**

Sécurité

Assurer la **robustesse** et la **résilience** aux conditions adverses, telles que le leurre et les cyber-attaques.

Explicabilité

Fournir des **explications compréhensibles et adaptées au contexte**.

Responsabilité

Se conformer aux **cadres éthiques, juridiques et réglementaires**

REF Thales/PCC - 2025/JM,30 – 13 juin 2025

Ethics

Recommendations from organizations like UNESCO and the OECD, or from EU high-level expert groups (HLEG)

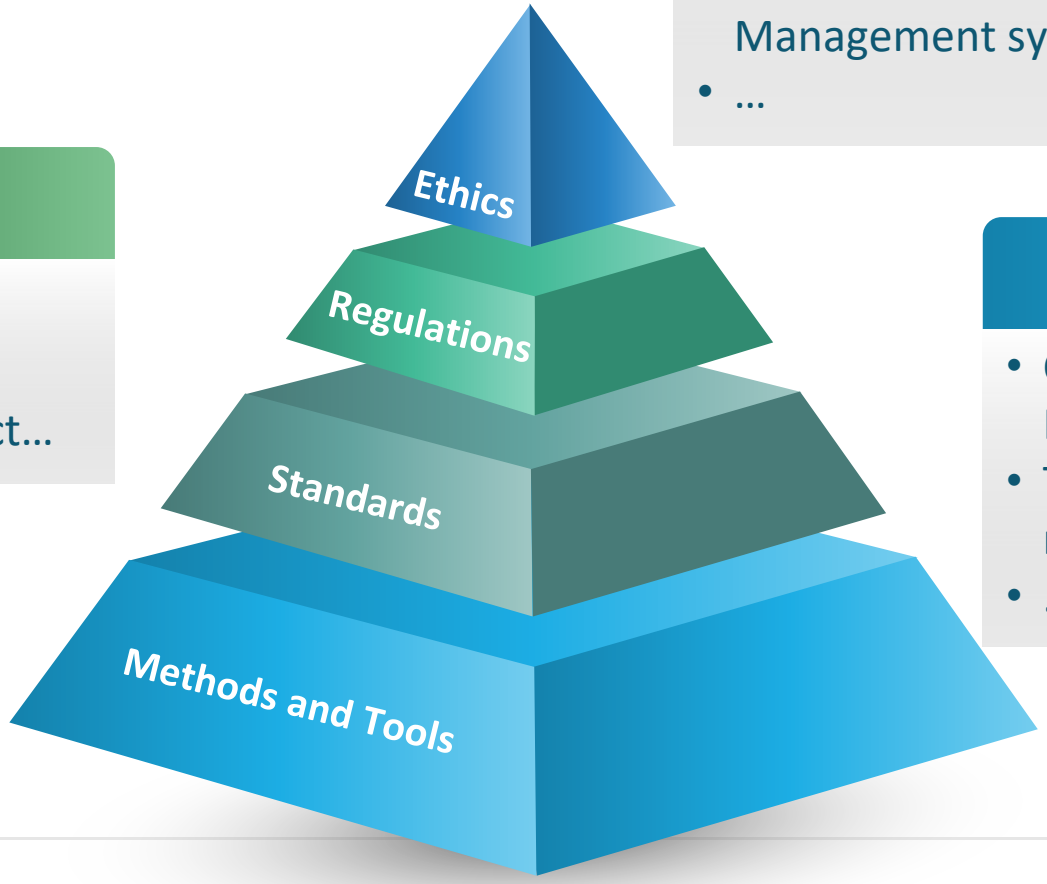
Standards

Glossary and technical requirements e.g.

- ISO/IEC 22989: AI concepts and terminology
- ISO 5338: the life cycle of AI systems based on ML
- ISO/IEC 23053: Framework for AI Systems Using ML
- ISO/IEC 42001: Information technology — AI — Management system
- …

Regulations

High Level, long-term requirements
e.g. European AI Act, Data Act…



Methods & Tools

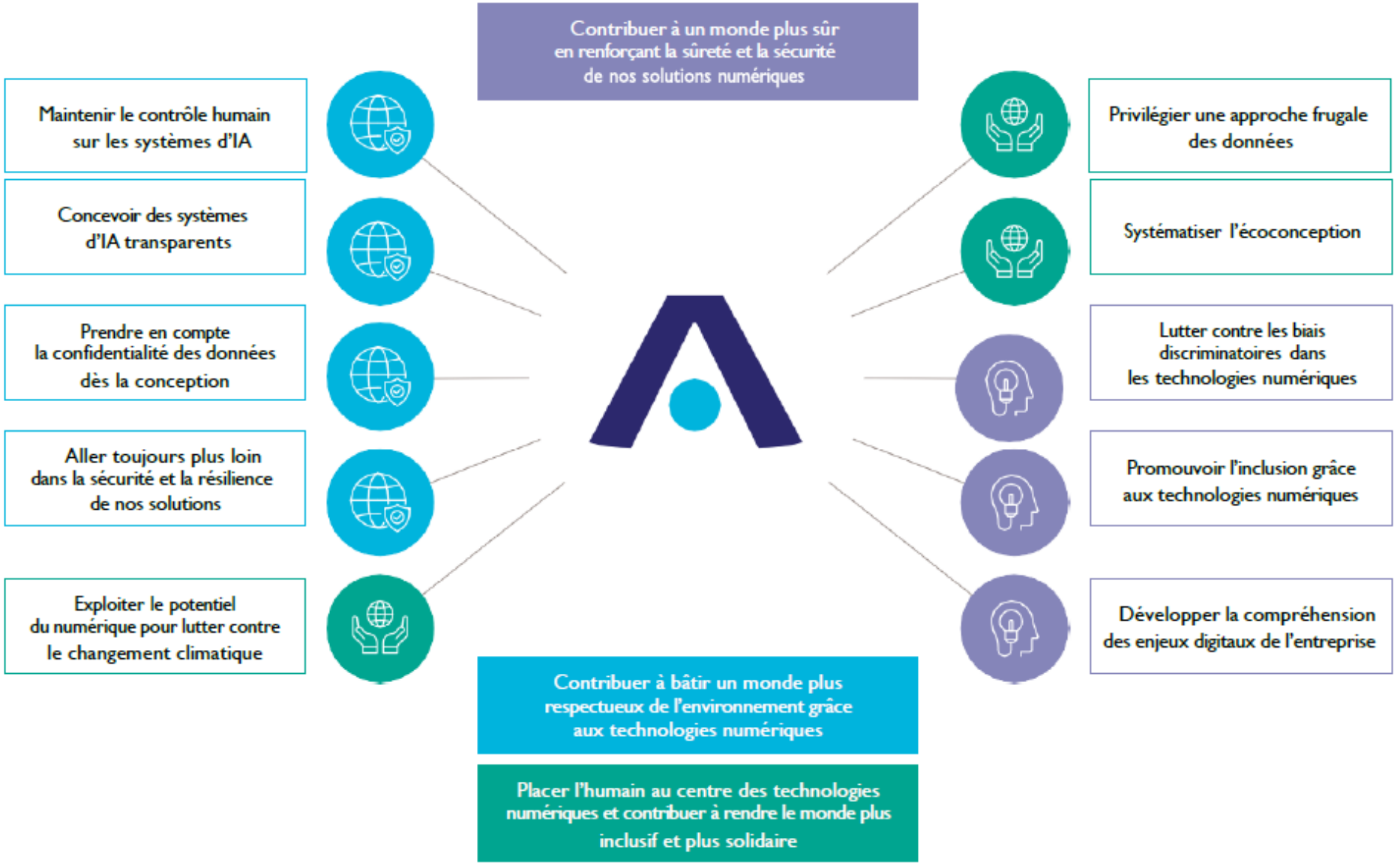
- Concepts of Design Assurance for Neural Networks – CoDANN [EASA]
- Tooled Confiance.ai End-to-End methodology
- …

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

REF Thales/PCC - 2025/JM,30 – 13 juin 2025

Les 10 principes de la charte éthique du numérique de Thales



REF Thales/PCC - 2025/JM,30 – 13 juin 2025

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

Les 8 technologies différentiantes de Thales en IA

IA Hybride
IA Connexionniste et Statistique
et IA Symbolique

IA frugale en donnée et énergie
Vers une IA à faible impact environnementale. Données simulées et synthétiques «Smart data» vs « Big data »

IA embarquée
Prise en compte des contraintes SWaP

AI Générative
LLM pour des systèmes critiques
IA générative de confiance,
fiable et Responsable

Apprentissage par
reforcement & Autonomie
Environnement de simulation
Digital Twins

Dialogue Homme-Machine
Interaction intuitive en fonction du contexte
IA auto-explicable

Intelligence Collaborative
Système multi-agents
IA distribuée

Ingénierie de l'IA de confiance
Conception, déploiement,
qualification, certification

REF Thales/PCC - 2025/JM,30 – 13 juin 2025

La confiance des systèmes critiques basés sur l'IA a un impact sur le cycle de vie global de l'ingénierie.

Nécessité de disposer des méthodes et outils d'ingénierie pour soutenir le cycle de vie global des systèmes critiques basés sur l'IA...

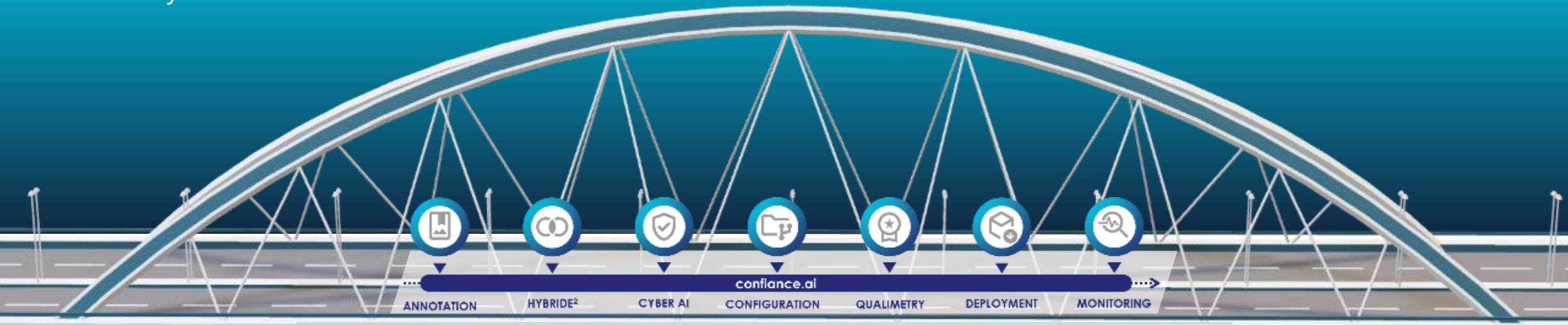
... Pour assurer la qualification/certification et la conformité avec les règlements et les normes

L'ingénierie de l'IA de confiance (cortAIx Labs & cortAIx Factory)

Des exigences et des spécifications
→ Exigences des parties prenantes
→ Spécifications du système et de l'IA/ML
→ Archis. système & AI/ML

Atelier AIOps

Au déploiement et à la maintenance
→ Monitoring
→ Vers la qualification et la certification



L'ingénierie de l'IA : une condition nécessaire pour déployer une IA digne de confiance

Operational Design Domain (ODD)

Validation

Vérification

Analyse opérationnelle de l'objectif visé (Intended Purpose)

Garantir la qualification et le respect des réglementations (ex, Assurance qualité) **et des normes (**ex, l'ARP6983 en aéronautique).

L'IA de confiance : condition nécessaire au déploiement de l'IA dans les systèmes critiques

Validité

Garantir qu'un système à base d'IA fait ce qu'il doit faire, tout ce qu'il doit faire et seulement ce qu'il doit faire

Transparence et explicabilité

Fournir des justifications et des explications compréhensibles et adaptées au contexte.

Responsabilité

Respecter les cadres éthiques, juridiques et réglementaires et être conformes aux standards

Sécurité et robustesse

Garantir la robustesse et la résilience aux conditions adverses, telles que le leurre et les cyber-attaques, mais aussi au mauvais usage

Gouvernance de la donnée

(RGPD et qualité de la donnée)

Fiabilité et sûreté

Contrôler le risque de défaillances inacceptables et d'insuffisances fonctionnelles

REF Thales/PCC - 2025/JM,30 – 13 juin 2025