

Stratégies d'industrialisation de l'IA de confiance pour les applications industrielles critiques

Conférence CentraleSupélec Santé

18 novembre 2025

Fabien Mangeant, Ex-Président du programme Confiance.ai, Group Chief Data & Al Officer Air Liquide

Quelles applications critiques de l'IA?



Sûreté Safety-critical Contrôle Qualité Pièces Sécuritaires



Suivi des Patients Care Plan



Avion Autonome



Financier
Business-critical

Gestion Automatique des Avis Client



Planification de la Production de Gaz



Système Bancaire



Support à la Décision

Optimisation

Système
Autonome 2



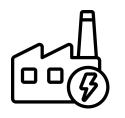
Quelles IAs ? Quels systèmes? Quels attributs de confiance?

ALGORITHMES Generative Al Unsupervised Supervised Learning Learning Reinforcement Learning Expert **Systems** Operational I

Research

SYSTEMES







ATTRIBUTS













Accountability Explainability Fairness



000

Human Oversight









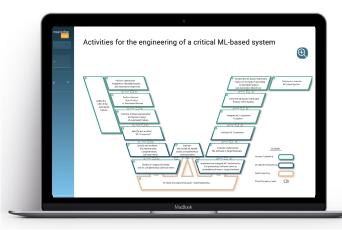
Des principes de gouvernance, un cadre réglementaire et des initiatives de standardisation très mouvants





Confiance.ai outcomes

- **50** Federated partners
- 13 Industrial use cases addressed
- 50+ Software components
- Guides and Methods
- 85 Scientific publications
- 9 Topics contributed to standards
- Participation in CEN-CENELEC & ISO JTC21 plenary sessions



A state of the art end-to-end engineering method



A catalog of guides and tools

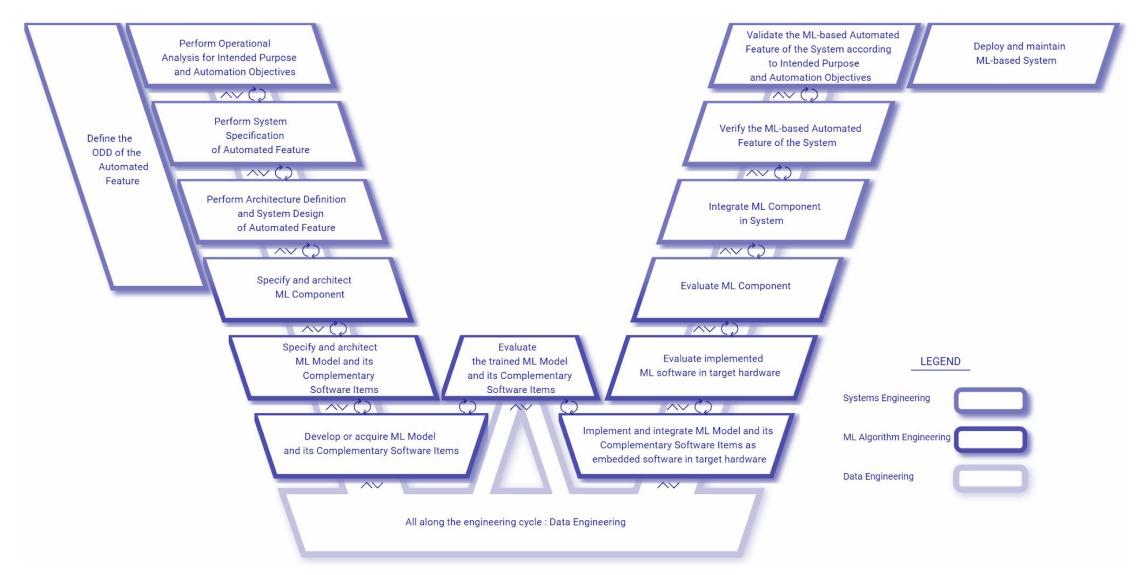


A strong industrial & scientific ecosystem

28/10/2025 C1 - Internal 5



The Body of Knowledge, and e2e method





Robustness

- Consistency
- ∠ Local Robustness

- Resilience
- Stability
- User Error Protection

Usability

- Learnability
- **⋈** Autonomy
- Accessibility

- Efficiency
- Memorability
- Data Usability

Effectiveness

- Appropriateness
- **区** Completeness
- Certainty

- Repeatability
- Representativeness
- ▼ Timeliness

Trustworthiness Characteristics

At system level

- ☑ Verifiability
- ☑ Understandability
- **X** Transparency
- ▼ Traceability

Human Agency

- ☑ Interpretability
- Explainability
- ☑ Comprehensibility
- ▼ Controllability

Dependability

- Availability
- Reliability
- Reproductibility

- ☑ Safety
- Security
- Maintainability

Respect for the rule of law

- Respect for fundamental rights
- Privacy
- Fairness

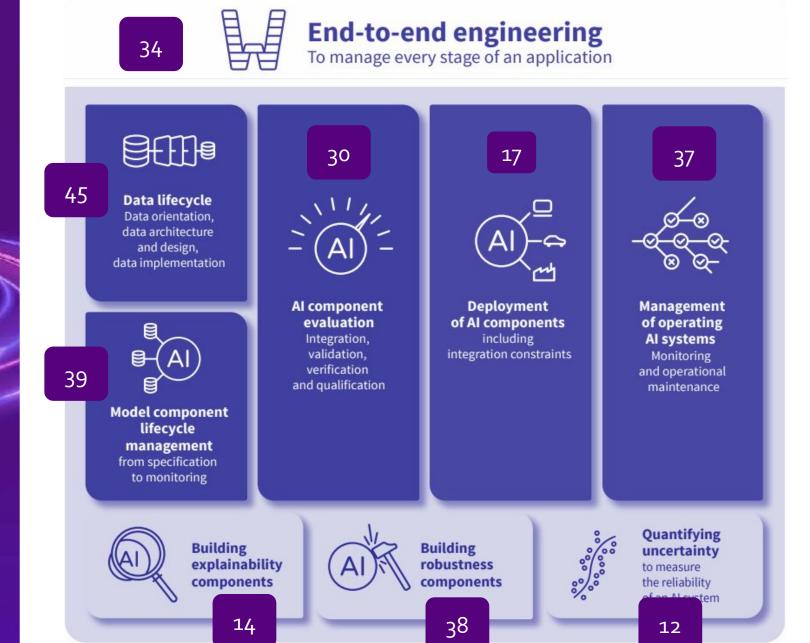
Human Oversight

- **E** Compliance
- Benevolence
- Accountability

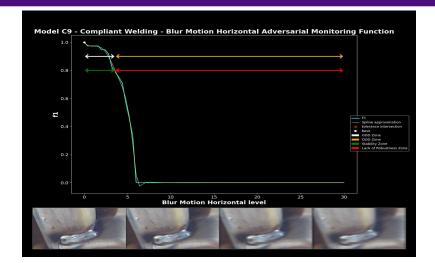


The Catalog

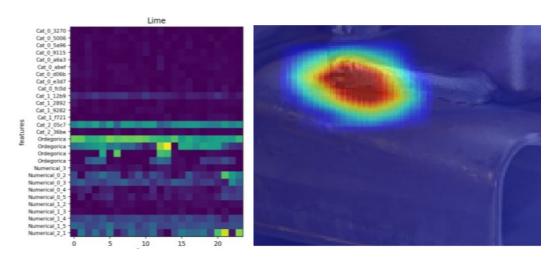
Functional sets



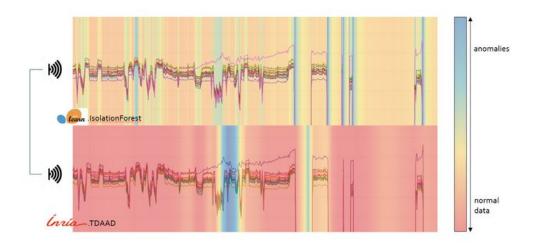
28/10/2025 C1 - Internal



MOODD - robustness analysis through image perturbation: analyze model performance degradation according to specific ODD condition



Kaa – a framework allowing to **explain** features that most influence the predictions of a model



Tadkit – a toolbox allowing to configure, compare, and combine a dozen of **anomaly detection** methods for time series to enhance decision-making, reduce risks or improve operational efficiency



Punce - a collection of state-of-the-art conformal prediction algorithms. Can be used with any predictive model to provide rigorous **uncertainty estimations**.



The European Trustworthy Al Association

Mission

a non-profit organization on a mission to empower Enterprises & Industry with state of the art, open source methodology and tools enabling the engineering of Al-based systems that can be trusted and comply with regulations.