



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 & AI Officer Air Liquide

Quelles applications critiques de l'IA?

Criticité
Sûreté
Safety-critical

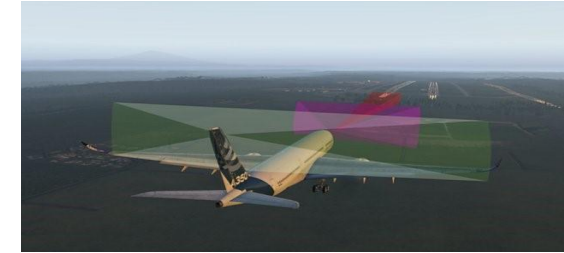
*Contrôle Qualité
Pièces Sécuritaires*



*Suivi des Patients
Care Plan*



Avion Autonome



*Gestion Automatique
des Avis Client*



*Planification de la
Production de Gaz*



Système Bancaire



Financier
Business-critical

*Support à la
Décision*

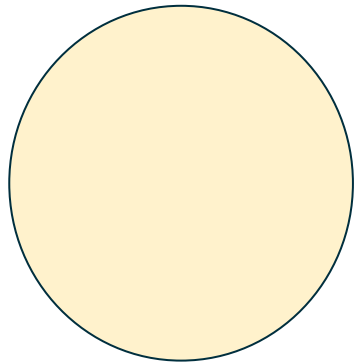
Optimisation

*Système
Autonome*

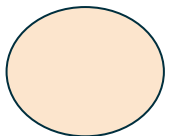
Autonomie

Quelles IAs ? Quels systèmes? Quels attributs de confiance?

ALGORITHMES



Supervised Learning



Expert Systems



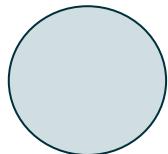
Generative AI



Unsupervised Learning

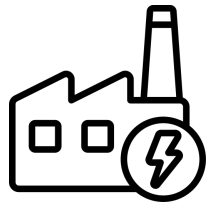
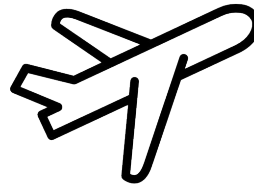


Reinforcement Learning



Operational Research

SYSTEMES



ATTRIBUTS



Privacy



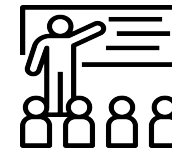
Security



Robustness



Accountability



Explainability



Fairness



Human Oversight

...



Quality



Safety



Maintainability

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

130+

Guides and Methods

85

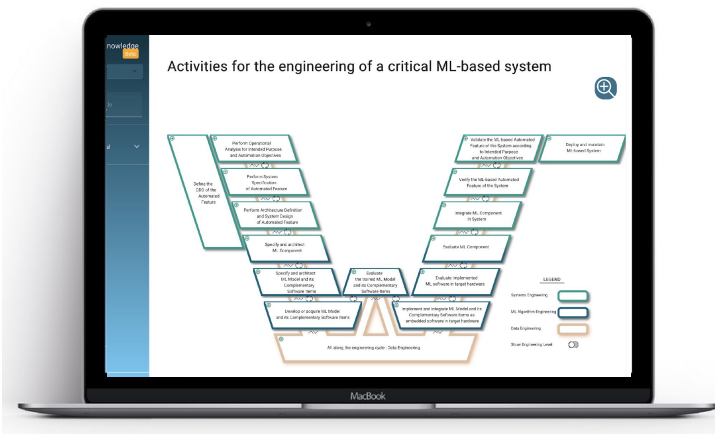
Scientific publications

9

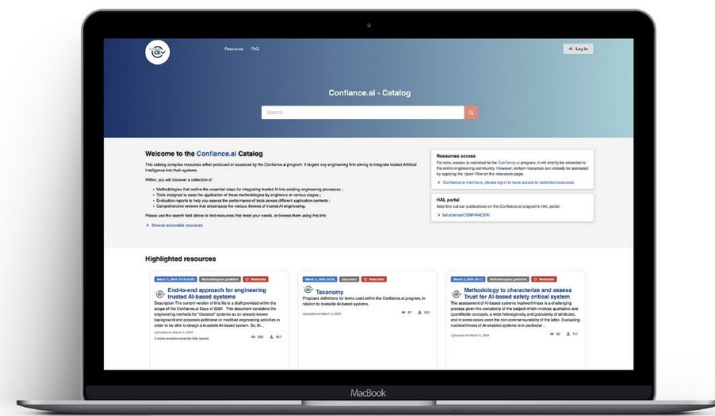
Topics contributed to standards

6

Participation in CEN-CENELEC & ISO JTC21 plenary sessions



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

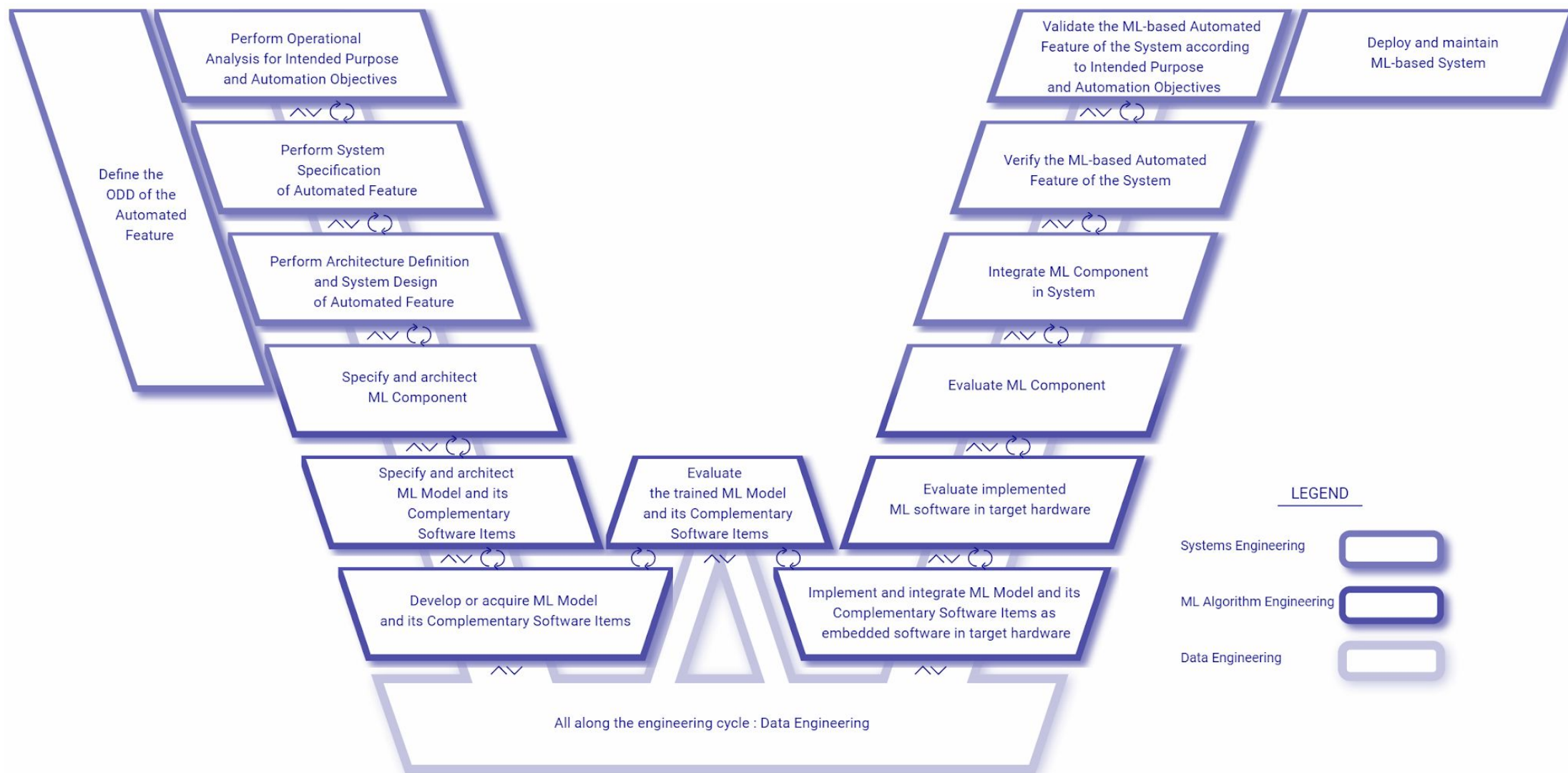


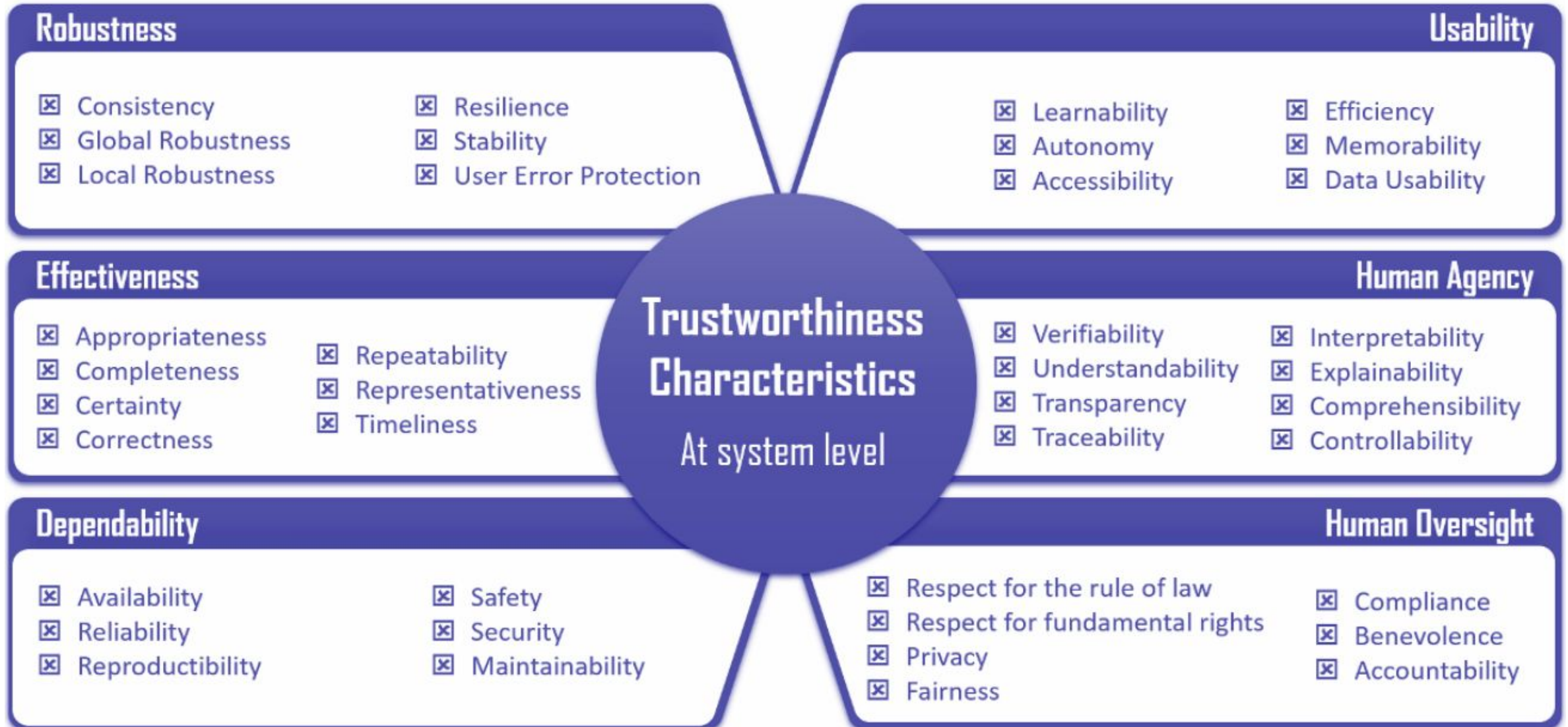
A catalog of guides and tools



A strong industrial & scientific ecosystem

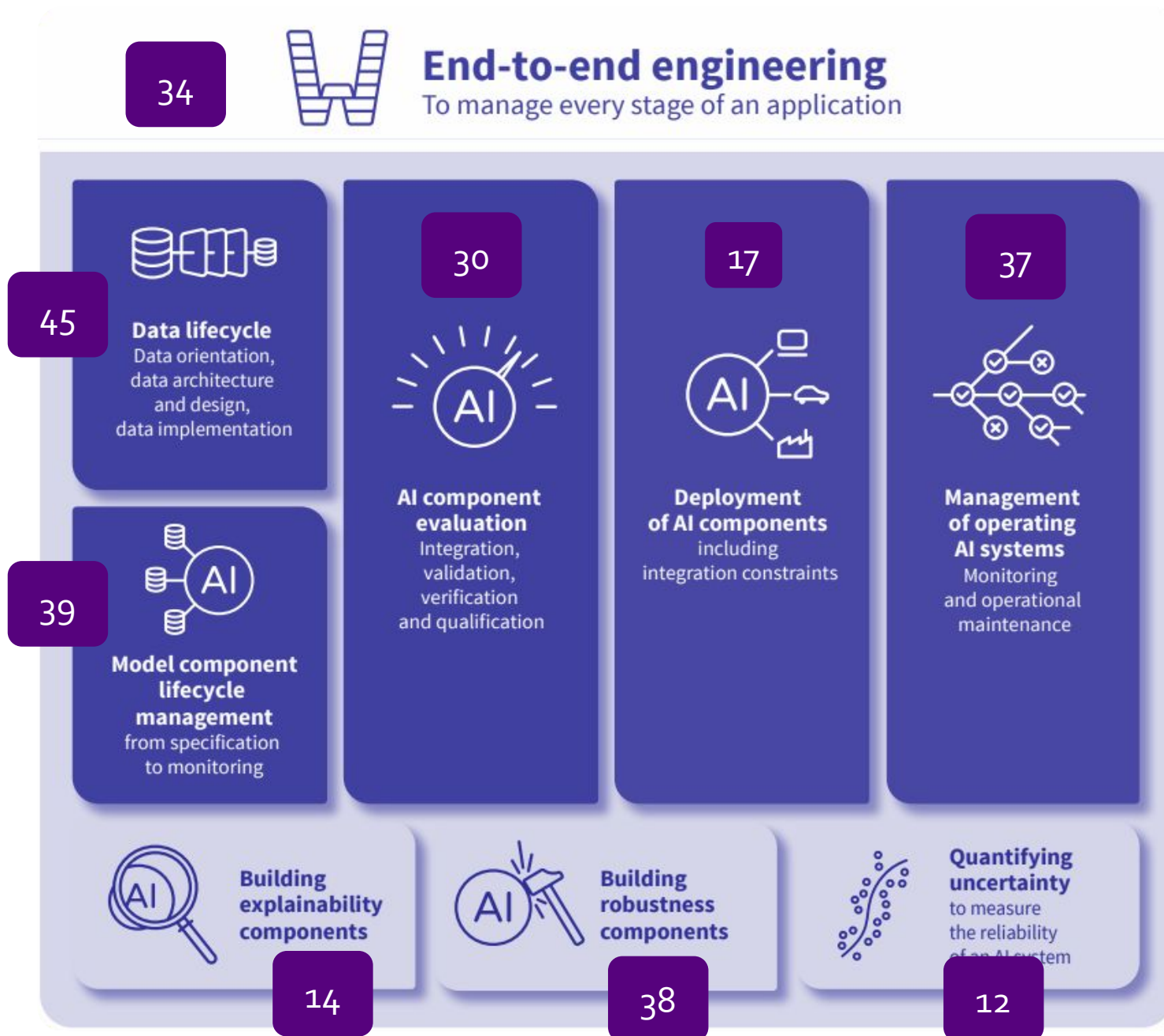
The Body of Knowledge, and e2e method





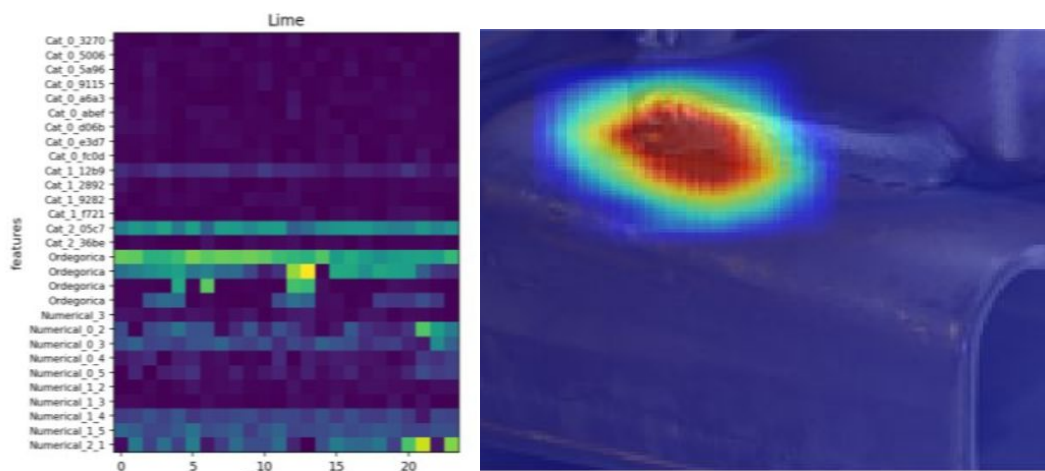
The Catalog

Functional sets

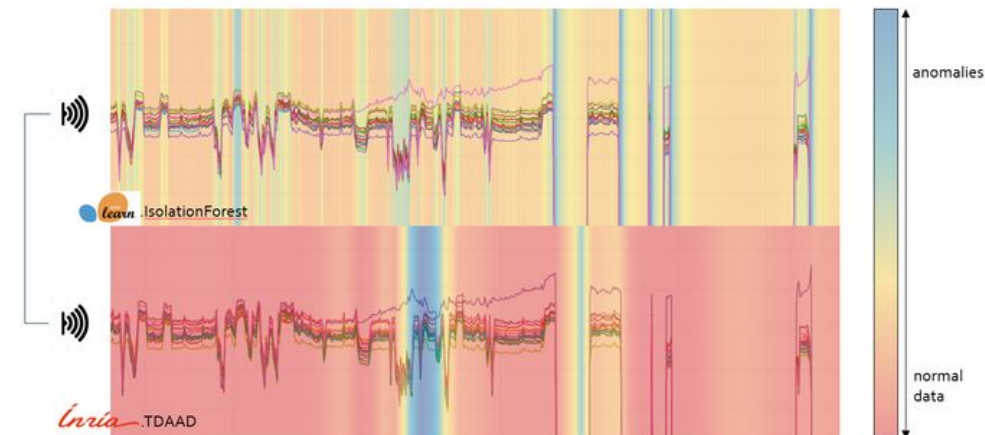




MOODD – **robustness** analysis through image perturbation : analyze model performance degradation according to specific ODD condition



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 AI 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 AI-based systems
that can be trusted and comply with regulations.