

# THALES

Building a future we can all trust



## IA de confiance :

Condition nécessaire pour le déploiement de l'IA dans les systèmes critiques

Juliette MATTIOLI



# AI for consumers and critical systems

## Safety-critical



- Failure may cause injury or death to human beings.

## Mission-critical

- Failure may result in the failure of some goal-directed activity.

## Business-critical

- Failure may result in the failure of the business using that system

	Consumers	Critical systems
Data	Open data, personal data, IoT	Critical data, truly anonymized data, critical sensors, HUMS, generated data, open data
Storage	Public clouds	Secured private & gvt clouds, embedded storage, public clouds
Computation	Mainly performed in the cloud, emergence of edge & fog computing	Cloud computing, edge computing, fog computing
Goal	Prediction of consumer activities, assistants, games	Prediction, optimization, decision (support), planning, collaborative systems
Security	Intrinsic security of the cloud	Certified, highest grade cyber security protection, trustable AI
Providers, Partners		

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of THALES - © 2021 THALES. All rights reserved.

# Scope: AI paradigm illustrated with some technics

**Communication:** Ability to understand language and communicate

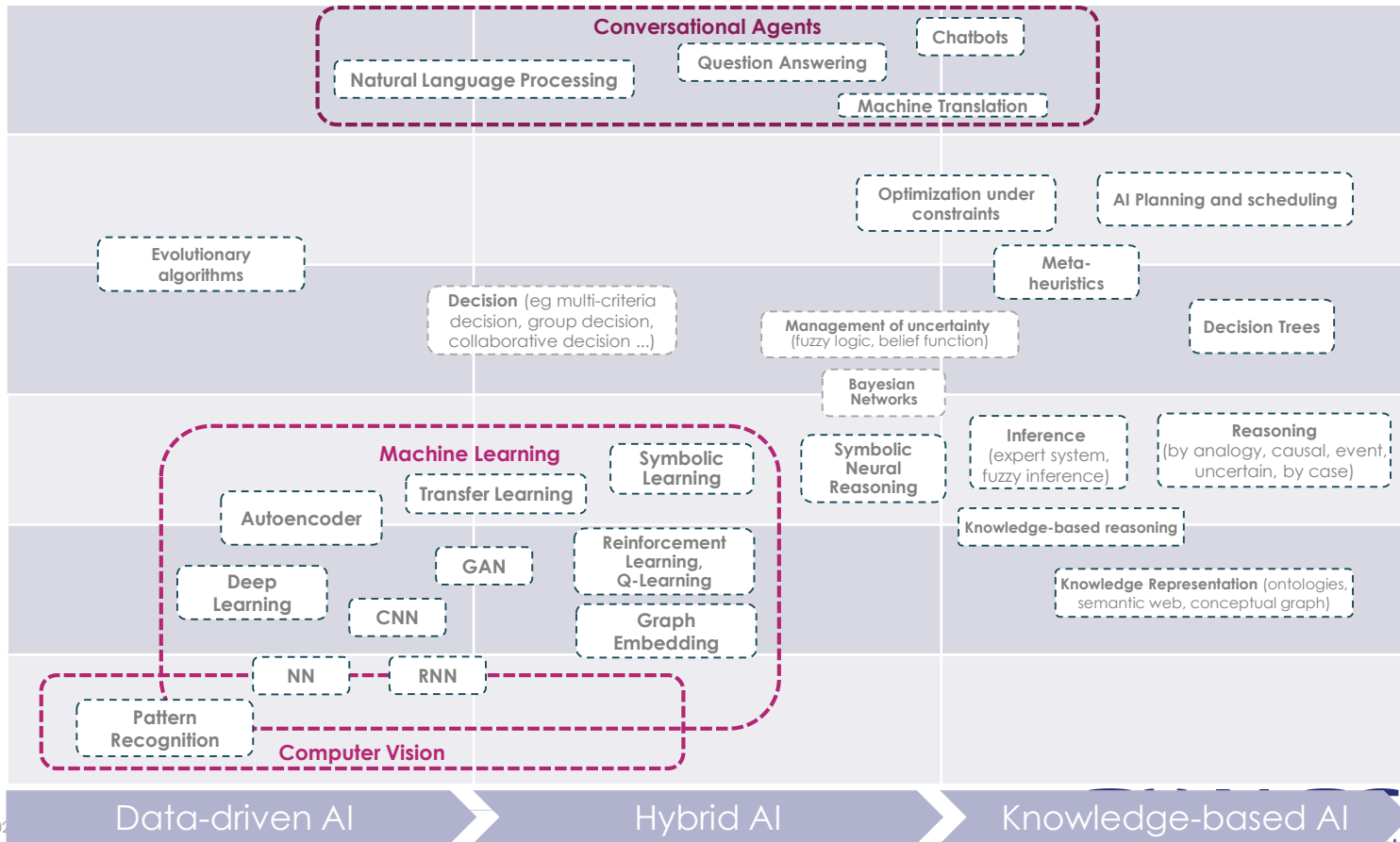
**Planning:** Capability of setting and achieving goals

**Decision:** Process of making choices among possible alternatives

**Reasoning:** the capability to solve problems

**Knowledge:** Ability to represent and understand the world

**Perception:** Ability to transform raw sensorial inputs (e.g., images, sounds, etc.) into usable information.



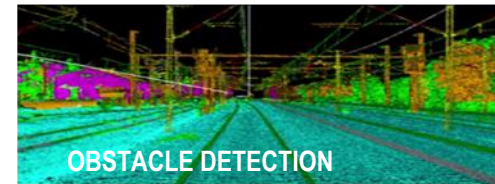
filed, published, translated, in any way, in whole or in part of THALES - © 2021 THALES. All rights reserved.

# AI Function: some use-cases in transportation domain



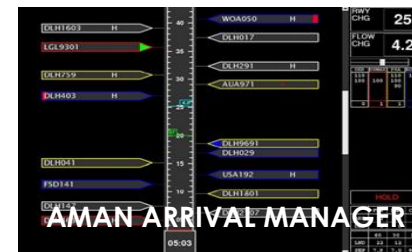
## Machine Learning (Neural Network, Deep Learning...)

- Autonomous train: Obstacle detection, train positioning, driver behavior classification
- Aerospace: Flight pattern learning
- Public transport network: Railway stations clustering, Airport congestion, Passengers' behavior analysis (violent behavior detection)
- HUMS: Anomaly detection



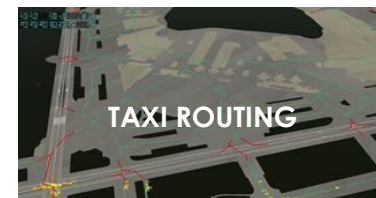
## Prediction (Convolutional Neural Network, Reinforcement Learning...)

- Public transport network: Passengers flow prediction and optimization, prediction of disruptive event, prediction of delays...
- Real-time accurate trajectory identification and anticipation of moving objects (aircrafts, ships, trains)
- Predictive maintenance for an airplane/train fleet (Remaining Useful Life)



## Decision making and Optimization (Genetic Algo, Constraint Solving, Path Planning, Multi-Criteria Decision Making...)

- Trajectory conflict resolution
- Train scheduling
- Sequencing for Delays Reduction: AMAN/DMAN
- Taxiing & Taxi Routing
- Aircraft fuel optimization



OPEN

# Critical AI-based system induced challenges

## Data & Knowledge



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

## Algo



- ✓ Specificifiability
- ✓ Traceability
- ✓ Correctness
- ✓ Accuracy
- ✓ Complexity
- ✓ Transparency
- ✓ Vulnerability mitigation (Robustness by design)

## Human-AI interaction



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

## Safety & Security



- ✓ Provability
- ✓ Verifiability (test)
- ✓ Robustness
- ✓ Resilience

## SW & SYS



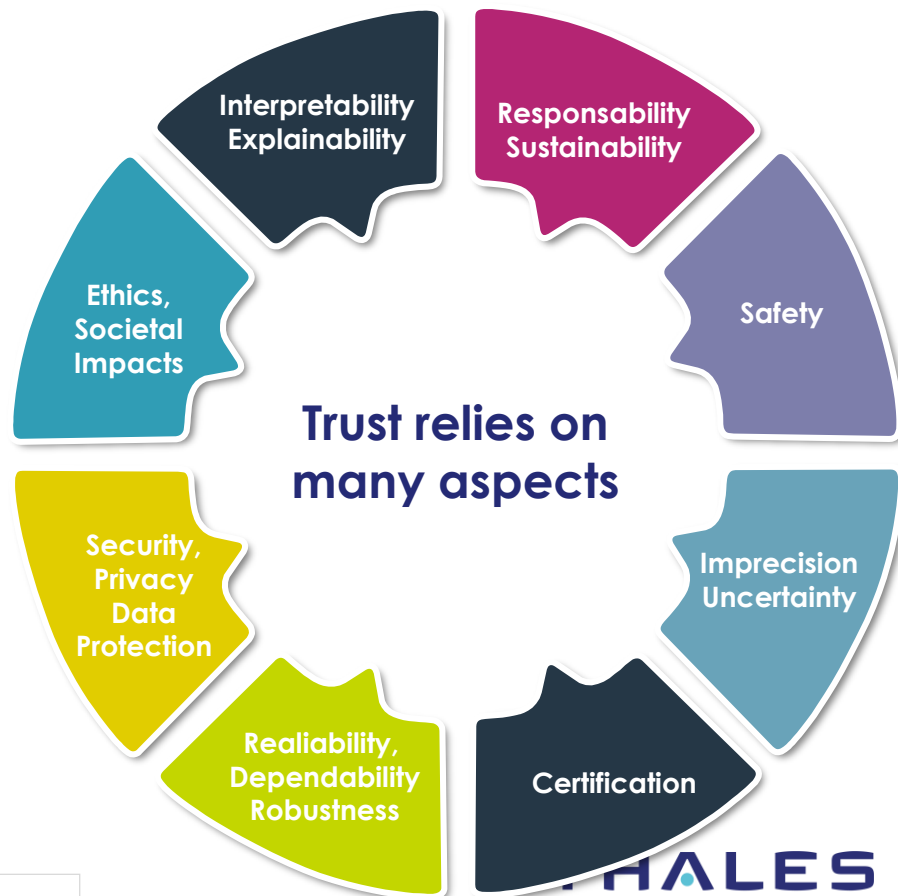
- ✓ Repeatability
- ✓ Performance
- ✓ Maintainability
- ✓ Auditability
- ✓ Monitorability

# Trustworthy AI Engineering

**Why:** to bridge the gap between AI PoCs and AI based critical system/solution

**How:** to support AI engineering processes and practices through methods, guidelines and interoperable tools during the overall system lifecycle by revisiting

- Trustworthy AI Taxonomy
- Algorithm Engineering
- Data and Knowledge Engineering
- Software and System Engineering
- Safety and security engineering
- Cognitive Engineering (human in the loop, ethics and other concerns)
- through the AI prism and safety & cyber-security issues



# Thales TrUE AI Strategy

## 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

# Trustworthy AI



## Security

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

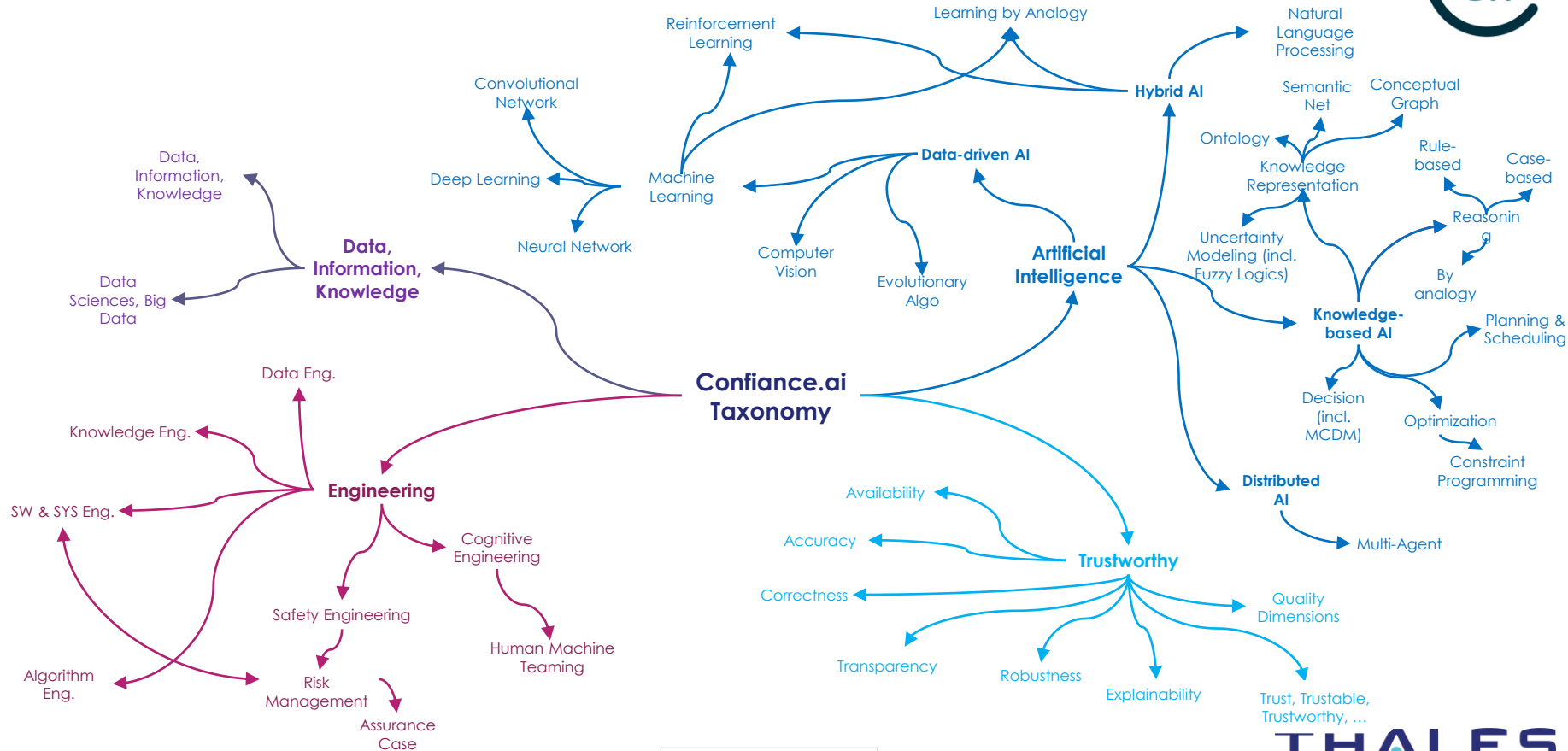
## Explainability

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

## Responsibility

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

# Trustworthy AI Taxonomy V1 (a 1st Draft – Sept 2021)



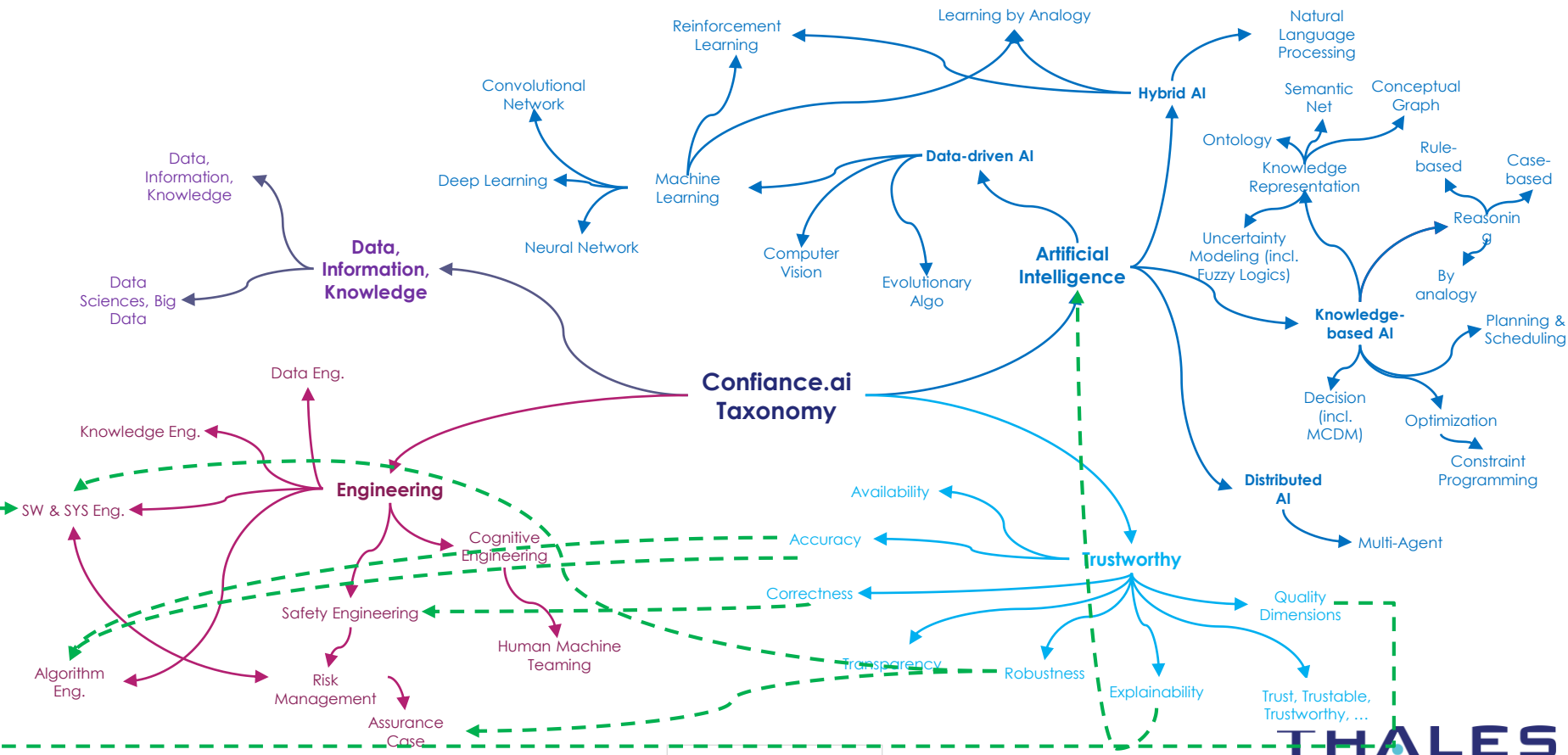
This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of THALES - © 2021 THALES. All rights reserved.

OPEN



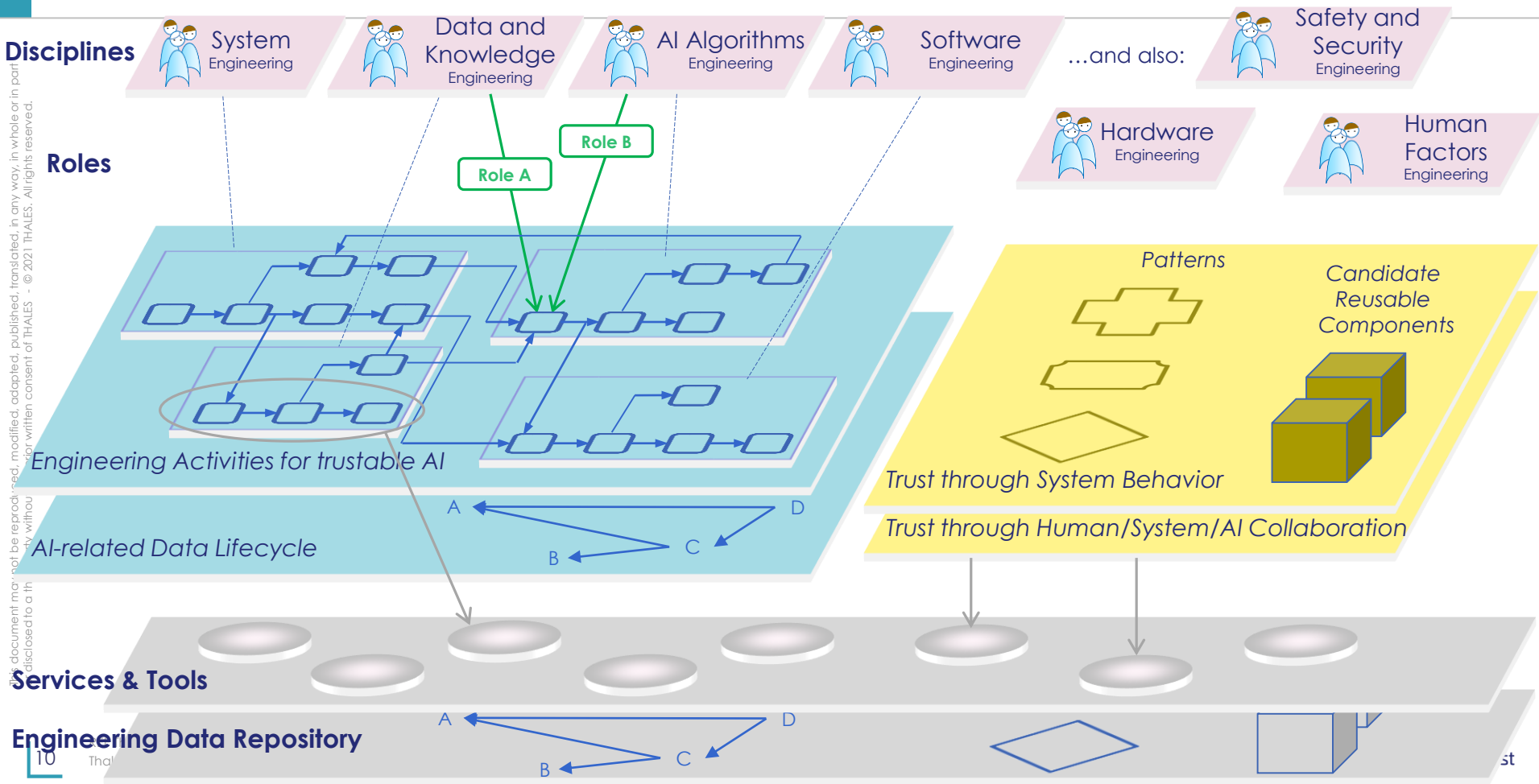
# More than a simple taxonomy... examples of trustworthy links

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part without the prior written consent of THALES - © 2021 THALES. All rights reserved.



# Critical AI-based system life cycle: a multi-disciplinary issue

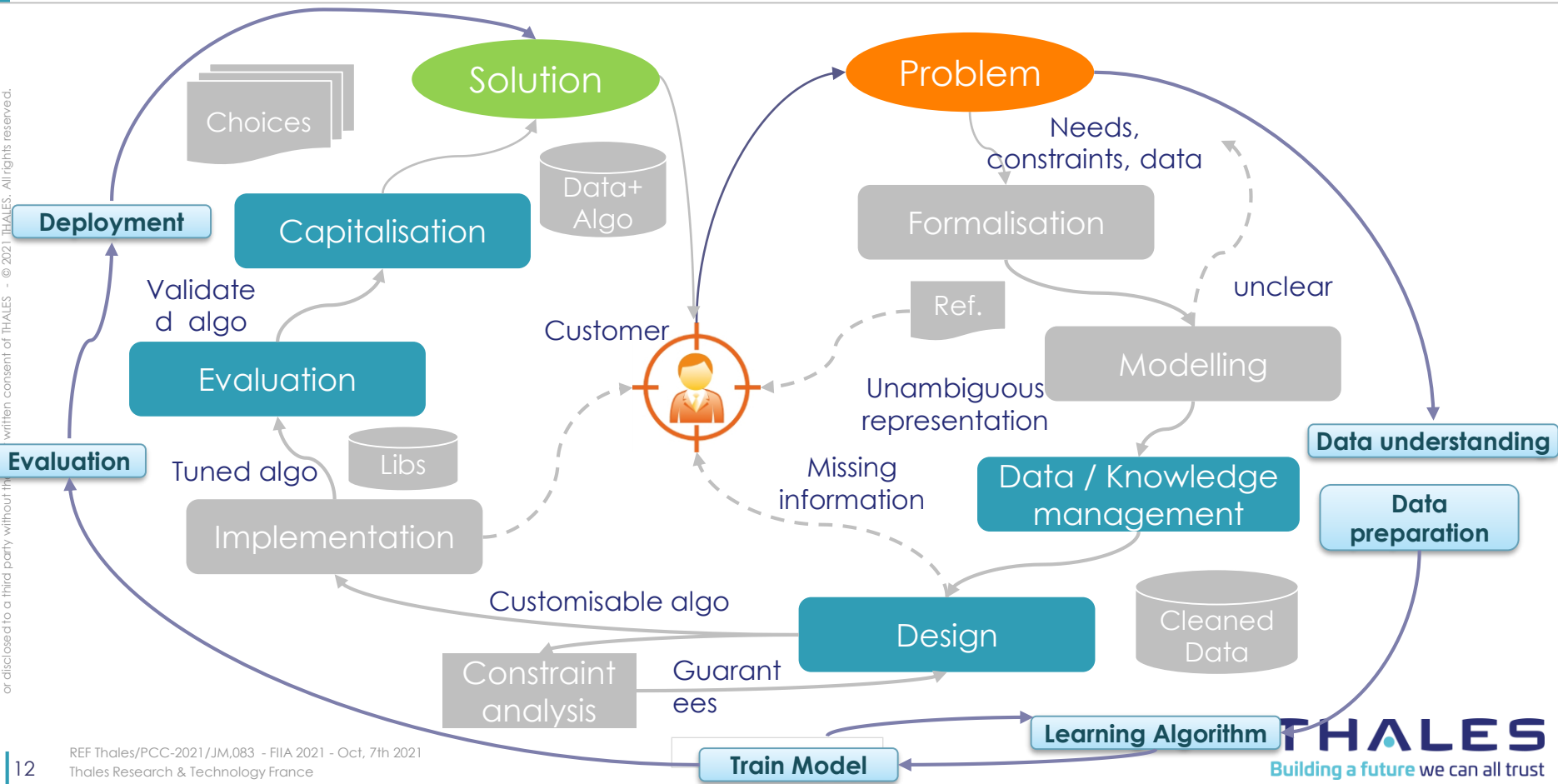
This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part, without the written consent of THALES - © 2021 THALES. All rights reserved.





# Mapping Algo Engineering steps within ML algorithm pipeline

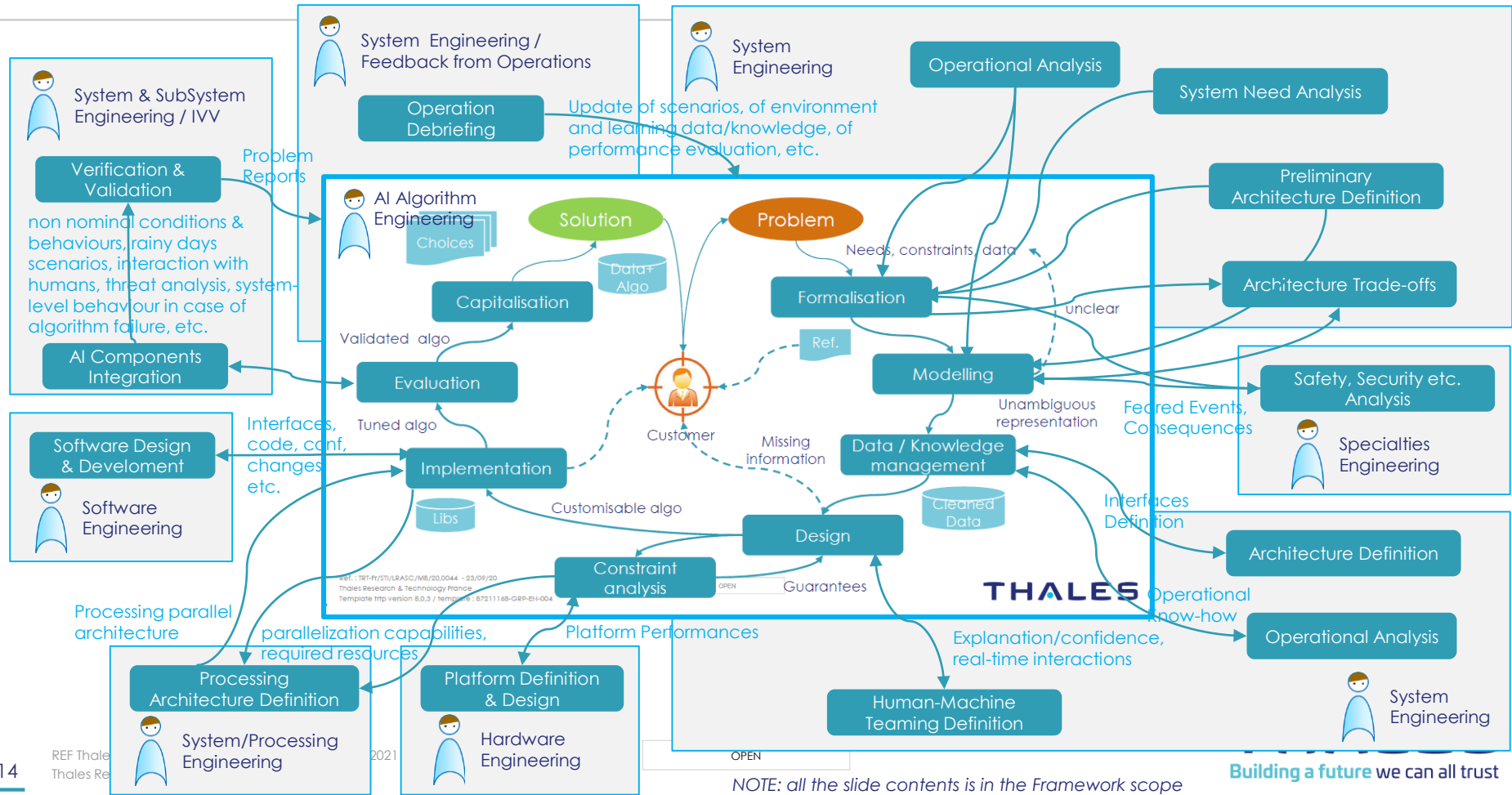
This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part without the written consent of THALES - © 2021 THALES. All rights reserved.





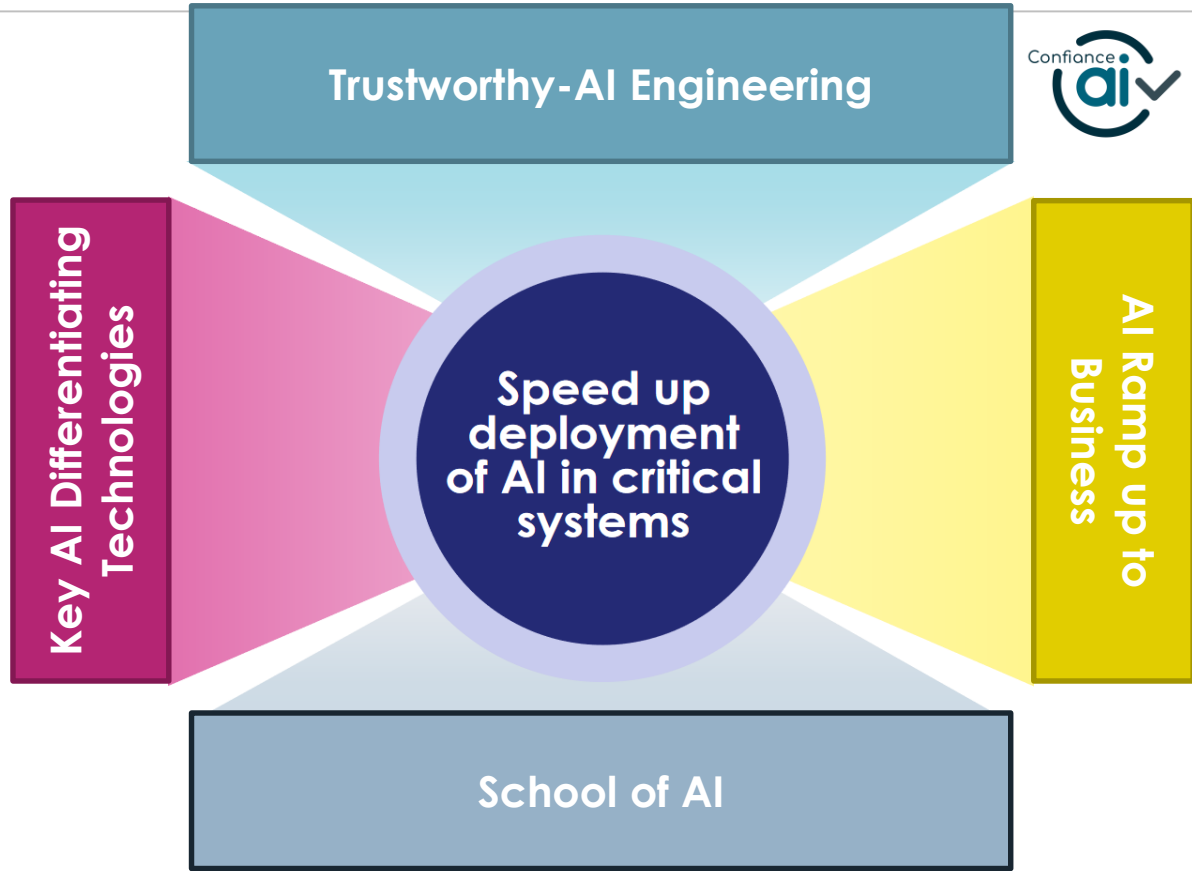
# Engineering Activities: Illustration Sample

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of THALES - © 2021 THALES. All rights reserved.



REF.: TRT-FY/STI/ARASC/MB/20.0044 - 23/09/20  
 Thales Research & Technology France  
 Template Imp version 8.0.3 / template: 87211168-GRP-EN-004

# Conclusion: Components of Thales TrUE AI Strategy



This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of THALES - © 2021 THALES. All rights reserved.