EC2 project: Processes and Methods

Systems Engineering report – v1

EC2 L4.1.2

EC2 L4.3.1

EC2 L4.4.1

EC2 L4.5.1

EC2 L4.6.1

EC2 L4.7.1

contact@confiance-ai.fr | www.confiance.ai

Specific information to be provided in ANR reports only

X	Progress report		Final report
From:	AAAA/MM/JJ	To:	AAAA/MM/JJ
Brief description of the project (reminder)			
Contractual description of the project:		~~nom du fichier~~	
Reference budget:		~~nom du fichier~~	

RECOMMANDATIONS

Strikethrough text is used to illustrate and explain the headings. It should be removed and replaced with non-strikethrough text. Similarly, this recommendations block should be removed, along with the statement "Specific information to be provided in ANR reports only", prior to dissemination of the document.

Document reference: EC2 L4.1.2 - L4.3.1 - L4.4.1 - L4.5.1 - L4.6.1 - L4.7.1

Contributors

	Name	Organisation	Role
Responsible for the deliverable	MATTIOLI, Juliette	Thales	EC2 Leader
Scientific responsible	XX	XX	XX
Co-authors	VOIRIN, Jean-Luc NOWODZIENSKI, Pierre BACLET, Julien ROBERT, Boris	Thales DMS FR Thales GTS FR IRT Saint Exupéry IRT Saint Exupéry	Research Engineers

Document control[2]-

Revision	Date	Commentary	Author
1.0	20-Dec-2021	First issue	XX
	XX	XX	XX
	XX	XX	XX
	XX	XX	XX
	XX	XX	XX

TABLE OF FIGURES

A. Introduction

A.1 Program context

The goal of **Confiance.ai** program is to address the multiples challenges that arise from the integration and the use of Artificial Intelligence processes in critical systems. The program will deliver an environment, the Trustworthy environment: a set of interoperable technological components, each one covering one or several of the specific challenges previously outlined. This environment will be articulated as instantiable tool chains, whose use will be defined in a set of methodological guidelines. It will allow the design, the development, the validation of Artificial Intelligence modules prior to their integration and support into critical systems.

A.2 EC2 project

A.2.1 General description

The objective of the EC2 project of the **Confiance.ai** program is to revisit the classical engineering disciplines (algorithmic engineering, software engineering and systems engineering) with regards to the effects induced by the use of AI in critical systems. A rigorous and interdisciplinary approach will formalize the design and the validation of these AI-based critical systems, allowing to guarantee “safe & secure” deployment and maintenance in operational condition, compatible with business usages.

To achieve this, the EC2 project will consolidate, if need be, the different methodological works of the other **Confiance.ai** projects (“EC3 – Characterization for AI trust”, “EC4 – Design for AI trust”, “EC5 – Data, information and knowledge engineering for AI trust”, “EC6 – VVQ strategy and systems certification” and “EC7 – embedded AI”.

The EC2 project will also commit to deal with the less human-system relationships and as well as the maintenance in operational and security conditions, and will be in perfect synergy with the EC1 project. Finally, the EC2 project will integrate the constraints induced by the regulation, the norms and standards with a strong connection to pillar 3, particularly thanks to the industrial members of the Confiance.ai program who signed the « Manifeste IA ».

A.2.2 Work packages and tasks

The EC2 project is broken down into four work packages, each work package being carried out through several tasks.

The present document addresses WP4: T1, T3, T4, T5, T6 and T7.

- WP1: “Taxonomy of Confiance.ai”
  - T1: “Definition of the glossary and the underlying concepts of trusted AI”
  - T2: “Definition of the professions and roles of the different underlying engineering disciplines for trusted AI”

- WP2: “Algorithmic engineering of trusted AI”
  - T1: “Critical analysis and fine identification of locks from the state of art about AI algorithmic engineering in a safe & secure context”
  - T2: “Definition of evaluation methods and KPIs of algorithmic properties”
  - T3: “Definition of evaluation methods and KPIs for the algorithmic engineering of trusted AI”
  - T4: “Methodological guide for the algorithmic engineering of trusted AI”
- WP3: “Data and knowledge engineering”
  - T1: “Critical analysis and fine identification of locks from the state of art about data engineering methods”
  - T2: “Methodology of data engineering for trusted AI”
  - T3: “Evaluation method and KPIs about the quality of datasets et knowledge basis for trust”
  - T4: “Methodology of knowledge engineering for trusted AI”
- **WP4: “Software and Systems Engineering for AI-based trusted systems”**
  - **T1: “State of art of evolutions of Systems Engineering principles, including MBSE, for insertion of AI”**
  - T2: “Assessment of the industrial situation of methods and processes used on the platforms of industrial partners”
  - **T3: “Operational Domain Definition”**
  - **T4: “Design, development and continuous integration methodology”**
  - **T5: “Methodology for taking into account human-system relationship”**
  - **T6: “IVVQ of a trusted AI-based system”**
  - **T7: “Methodology of maintenance in operational condition, safety and (cyber-) security”**
  - T8: “Methodological guide for Systems Engineering of trusted AI”

A.3 Addressed deliverables of the project

The table below indicates which chapters of the present document corresponds to the deliverables planned in EC2’s contractual descriptive document.

Deliverable in EC2 contractual descriptive document	Corresponding chapter(s) in the present document
EC2 L4.1.2: “Roadmap of Systems Engineering (including MBSE) method for AI insertion”	§C.11 *Viewpoint “Engineering Environment Capabilities”* as well as the other sub-chapters of §C *Definition of the framework through architecture viewpoints*
EC2 L4.3.1: “Chapter of the methodological guideline for trusted AI Systems Engineering, dedicated to operational domain definition (V1)”	§C.5 *Viewpoint “Risk on Trust due to System in Operation”*
EC2 L4.4.1: “Chapter of the methodological guideline for trusted AI Systems Engineering,	§C.2 *Viewpoint “Engineering Activities for trustable AI”* §C.3 *Viewpoint “AI-related Data Lifecycle”*

Deliverable in EC2 contractual descriptive document	Corresponding chapter(s) in the present document
dedicated to design, development and continuous integration methodology (V1)"	§C.8 *Viewpoint "Integration of AI functions"* §C.10 *Viewpoint "Engineering Lifecycle Management"*
EC2 L4.5.1: "Chapter of the methodological guideline for trusted AI Systems Engineering, dedicated to the methodology for taking into account human-system relationship (V1)"	§C.7 *Viewpoint "Trust through Human/System/AI Collaboration"*
EC2 L4.6.1: "Chapter of the methodological guideline for trusted AI Systems Engineering, dedicated to the IVVQ of a AI-based system (V1)"	§C.2 *Viewpoint "Engineering Activities for trustable AI"* §C.4 *Viewpoint "Risk on Trust due to Engineering"* §C.8 *Viewpoint "Integration of AI functions"*
EC2 L4.7.1: "Chapter of the methodological guideline for trusted AI Systems Engineering, dedicated to the methodology for in-service support, safety and (cyber-) security after deployment (V1)"	§C.5 *Viewpoint "Risk on Trust due to System in Operation"* §C.6 *Viewpoint "Trust through System Behavior"* §C.9 *Viewpoint "Trust on Performance, Safety, Security"*

A.4 Glossary of terms and acronyms

Refer to « EC2 L1.1.1 - 1st version of Confiance.ai Taxonomy ».

B. General context of the framework

B.1 Proposed approach

The present document proposes to formalize part of information collected and elaborated by projects, into a “Trustable AI Engineering Definition Framework”, in order to:

- contribute to guide EC2 Analysis Activities,
- check, relate, consolidate and capitalize collected/elaborated information, in an incremental and evolutionary manner,
- give efficient means for other projects to explore and exploit these Information
- constitute a need definition input for EC1 workbench, including justification and traceability with engineering needs,
- constitute a major, digitized deliverable of EC2, to be used as a reference and guidance about **Confiance.ai**’s Processes, Methods, Practices for future users of **Confiance.ai**-delivered Workbench.

The approach consists in:

- analyzing engineering context, constraints, activities, data, lifecycle, etc., concurrently under different Viewpoints, in order to build a global, comprehensive understanding, each viewpoint enriching others in an iterative and incremental, multi-viewpoint analysis ;
- formalizing and analyzing correspondence rules, relationships between viewpoints, in order to enrich and consolidate their contents and coherency. *(As defined in ISO 42010: "Correspondences and correspondence rules are used to express and enforce architecture relations such as composition, refinement, consistency, traceability, dependency, constraint and obligation.")*

B.2 Analysis Drivers & Contents

The viewpoints described in §C have been built by:

- considering existing engineering practices and related activities:
  - AI algorithms engineering,
  - data and knowledge engineering,
  - system and software, hardware engineering,
  - human factors, cognition and knowledge engineering (human, social and societal levels),
  - safety (certification) and security engineering,
  - sustainable concerns (e.g. energy consumption).
- characterizing, prioritizing and filtering them according to two drivers:
  - specificities of AI,
  - trust-related concerns.
- identifying joint activities, and mutual impact of engineering decisions

- basing the analysis on different Use Cases, e.g. :
  - starting with “theoretical” UCs (such as “autonomy”, “sensor data analysis”, “symbolic decision aids” and “operational maintenance”),
  - then revisiting analysis by confronting it to use cases and outputs of other partners/projects.

B.3 Possible Analysis Outputs

This approach is expected to produce:

- a global & encompassing knowledge base of engineering, describing & linking:
  - Engineering Concerns, associated Practices, and links with processes,
  - Engineering Activities (and beyond), focusing on their specificities to address trustable AI,
  - Engineering and lifecycle data, involved in the former ;
- a description of applicative and generic services of the engineering environment (e.g. data transformation/confrontation/coherency, impact analysis, activity analysis, etc.) ;
- recommendations on solution architectures, to support former capabilities:
  - generic patterns or notional components (e.g. monitoring & control, decision analysis…),
  - recommended functional chains and operational scenarios,
  - possibly, a Product Line vision (for example at algorithm, software and system levels).

B.4 Role of this framework in Confiance.ai program

To ensure that AI-based systems will possess the necessary Trust Properties, it is required to elaborate specific strategies for System Development Activities (elaborated by **Confiance.ai** EC2) and V&V Activities (elaborated by **Confiance.ai** EC6).

These activities will use a chain of elementary Methods and Tools (specified/recommended by **Confiance.ai** EC3-EC4-EC5-EC7).

The System Development Activities (as elaborated by EC2) will produce Engineering Items. The V&V Activities (as elaborated by EC6) will produce the Evidences that these Engineering Items possess the Trust Properties.

The Trust Environment (produced by **Confiance.ai** EC1) will orchestrate these System Development Activities and V&V Activities, and will store or reference the produced Engineering Items and Trust Evidences.

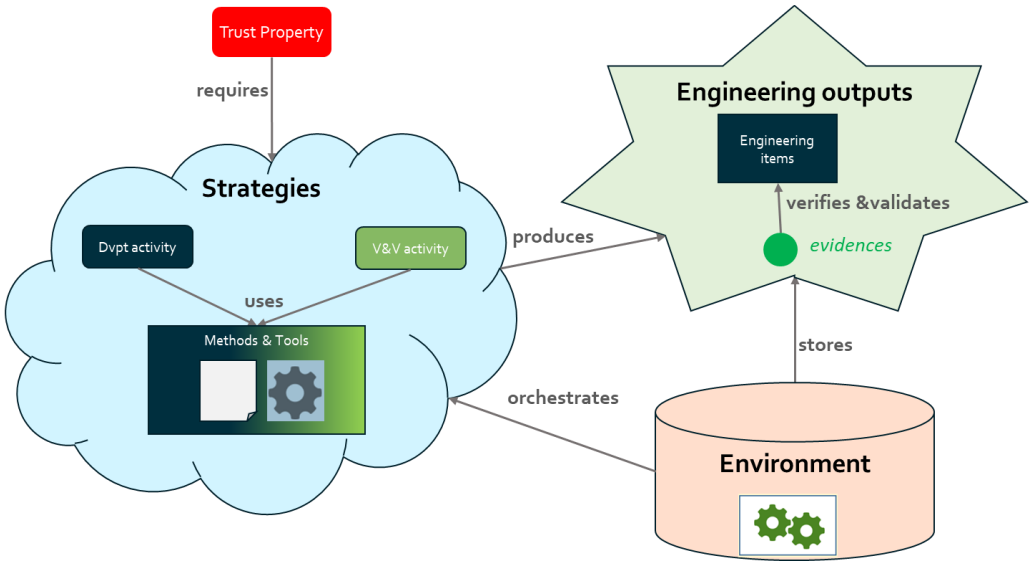

Figure 1: EC1-EC2-EC6 convergence

The “Trustable AI Engineering Definition Framework” proposed by EC2 in the present document will allow to elaborate these strategies for System Development Activities and, in collaboration with EC6, also these strategies for V&V activities. After having been communicated to EC1, it will contribute to the specification of this Trust Environment.

C. Definition of the framework through architecture viewpoints

C.1 Introduction

Each following chapter describes one the viewpoints resulting from the analysis described in §B.

Each viewpoint and each view (i.e. each snapshot of the « Trustable AI Engineering Definition Framework » general picture, according to a given viewpoint) has been modeled in Capella. At the end of the project, the resulting Capella model(s) may be the digitized deliverable described in §B.1.

This is still a work in progress. The definitions of each viewpoint may evolve during the **Confiance.ai** program and additional viewpoints may be created. Each view is just provided as a first draft of illustration for each viewpoint. The point of the present document is mainly to expose the approach that is currently started and that will be refined in the course of the EC2 project.

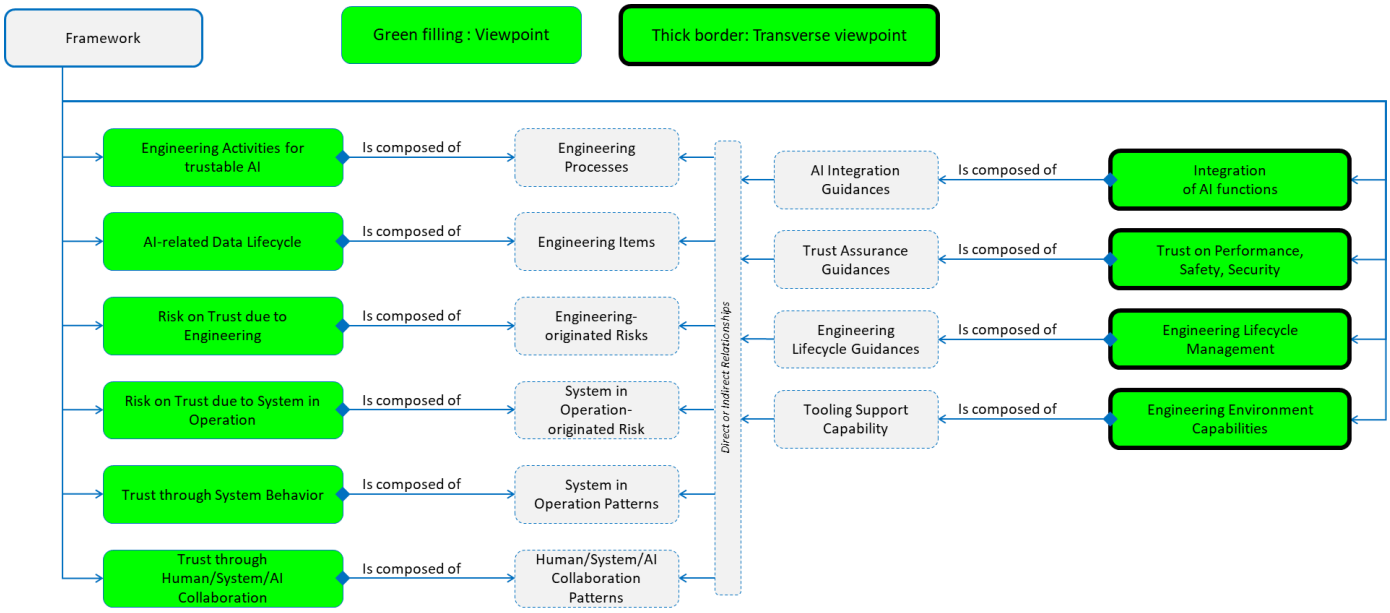

Figure 2: General structure of the framework

C.2 Viewpoint “Engineering Activities for trustable AI”

C.2.1 Purpose & contents

- Define the tasks to perform so as to specify, design, produce, deploy and operate an appropriate & trustable solution to a well understood need, involving AI techniques.

C.2.2 Justification

- Activities, roles and interactions/exchanges are the basis to describe methods and guidelines for a trustable AI enabled engineering.

C.2.3 Elaboration approach

- Formalize the end-to-end process to design an AI-empowered component, from need analysis to operations and feedback.
- Identify which engineering activities (Sys, SW, Specialties…) have to deal with AI and trust specificities, formalize their resulting adaptations.
- Detail how these processes & engineering activities interact, what they exchange, how they adapt and constrain each other.

C.2.4 Concepts and meta-model

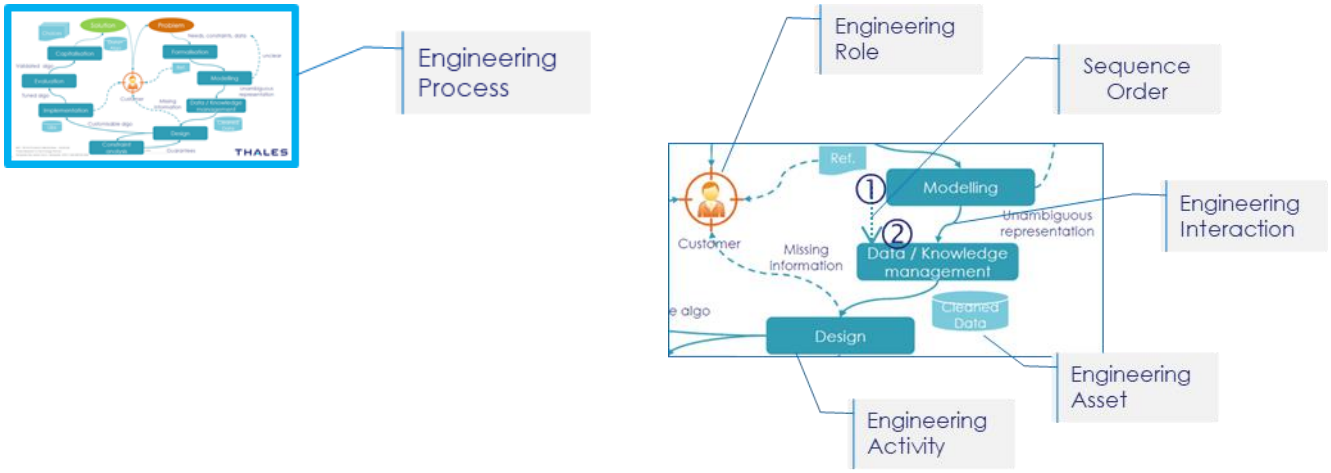

Figure 3: Concepts used in the "Engineering Activities for trustable AI" viewpoint

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

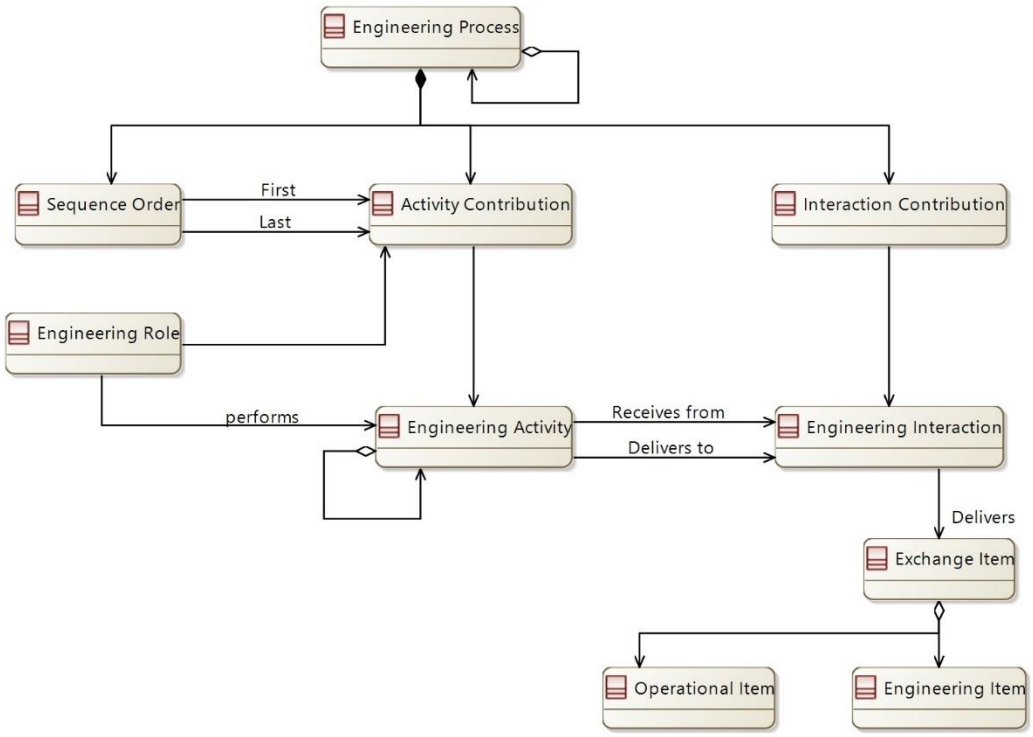

Figure 4: Meta-model of the "Engineering Activities for trustable AI" viewpoint

NOTE:

- "Engineering Activities" are described according to what they produce, what they need for so, and which other activities can deliver inputs to them, or can use their outputs (as defined by "Engineering Interactions").
- "Activity Contributions" describe the way "Engineering Activities" and "Engineering Interactions" contribute to an identified "Engineering Process", and their specificities in this particular context.

C.2.5 Illustration

"Engineering Activities for trustable AI" views are captured in the Operational Analysis (OA) perspective of Capella.

Mapping of the Viewpoint's concepts to Capella objects:

- "Engineering Process" maps to Capella OA's "Operational Process".
- "Sequence Order" maps to Capella's "Sequence Link".
- "Activity Contribution" maps to Capella OA's "Operational Process Involvement Operational Activity".
- "Interaction Contribution" maps to Capella OA's "Operational Process Involvement Link".

- “Engineering Role” maps to Capella OA’s “Operational Entity” or “Role”.
- “Engineering Activity” maps to Capella OA’s “Operational Activity”.
- “Engineering Interaction” maps to Capella OA’s “Operational Interaction”.
- “Exchange Item” maps to Capella’s “Exchange Item”.
- “Engineering Item” and “Operational Item” map to Capella’s Classes involved as “Exchange Item Element” of an “Exchange Item”.

The diagram below “Detailed View of AI Algorithm Development Activities” is only a first draft intended to illustrate a use of the concepts described in §C.2.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

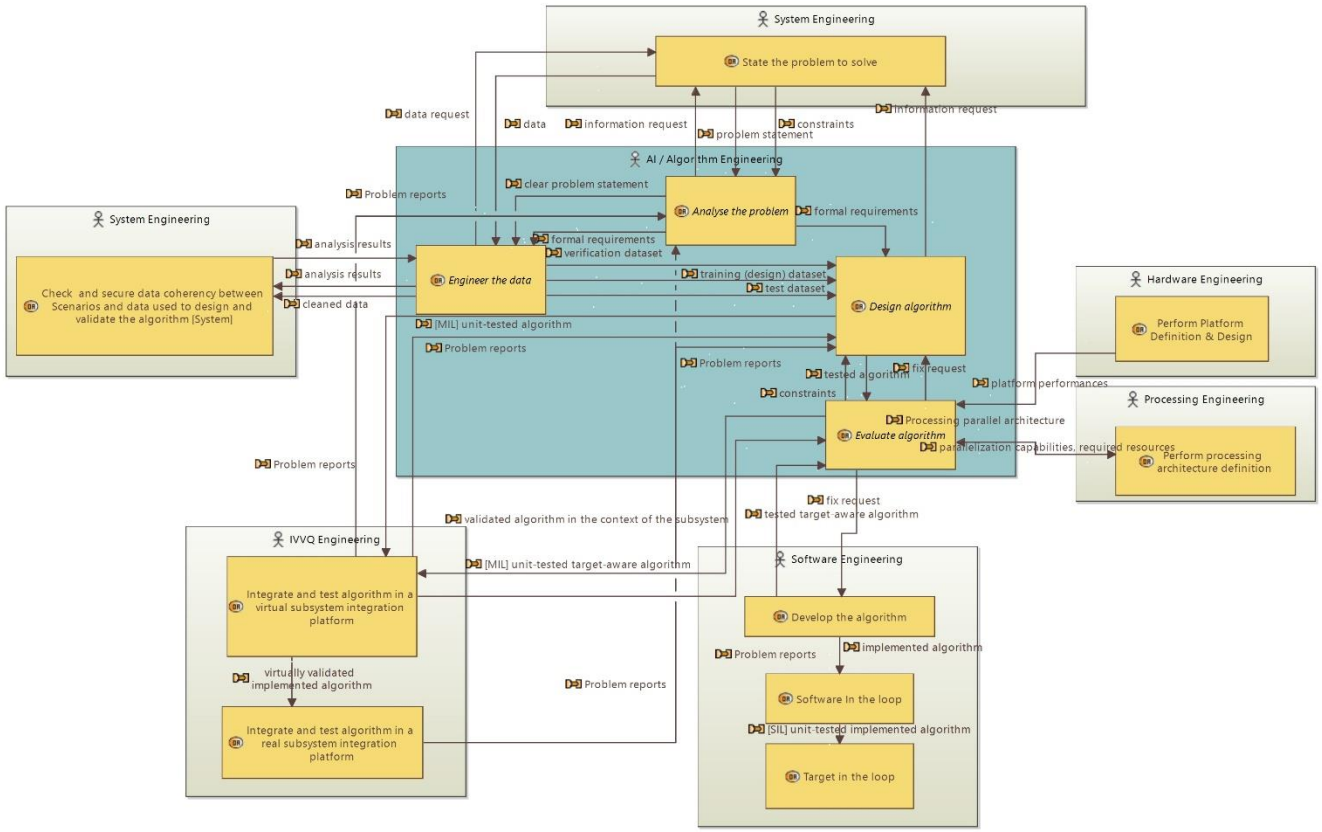

Figure 5: First draft of the view “Detailed View of AI Algorithm Development Activities”

C.3 Viewpoint “AI-related Data Lifecycle”

C.3.1 Purpose & contents

- Identify major data required/produced by AI engineering, when they are produced/used, and how they evolve during time.

C.3.2 Justification

- Most of these data will be part of interfaces between AI and other engineering or operation, and will require actions from most engineering stakeholders.
- They are likely to evolve during time, which may deeply constrain engineering.
- They are key to explain how engineering activities will perform.

C.3.3 Elaboration approach

- List data for AI; list engineering data; list operation/exploitation data
- Analyze how they are related, and/or built from each other.
- Consider their evolution according to context and environment changes.
- Link to appropriate engineering activities, as activities inputs/outputs.

C.3.4 Concepts and meta-model

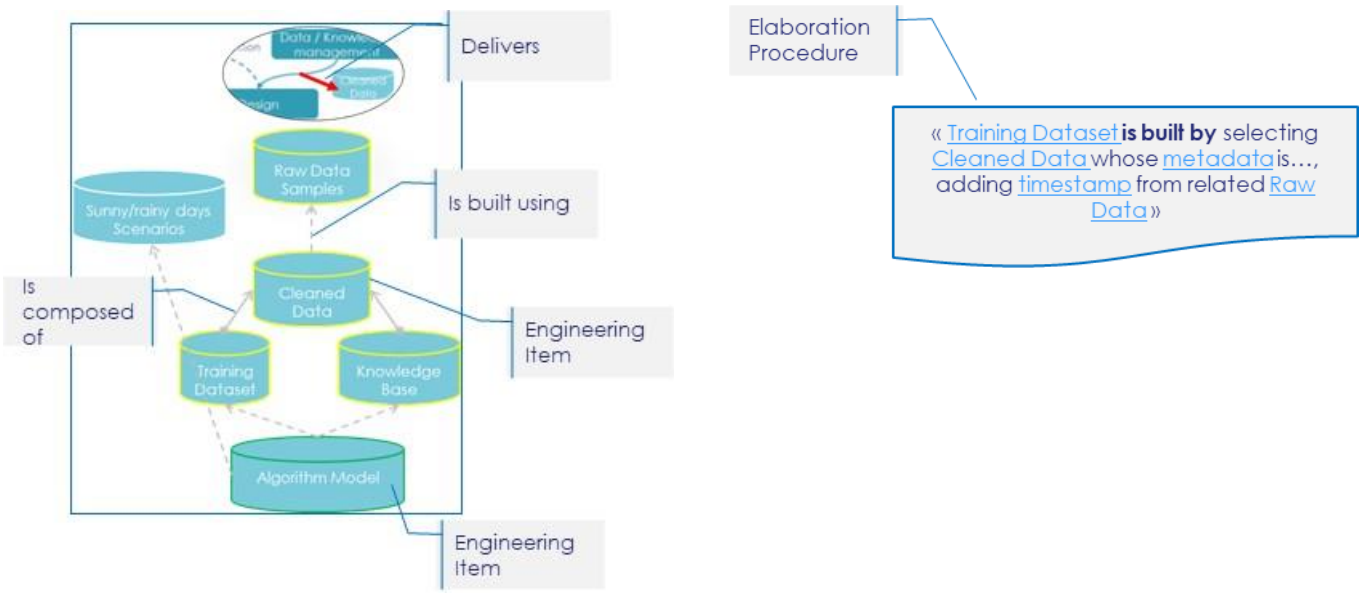

Figure 6: Concepts used in the "AI-related Data Lifecycle" viewpoint

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

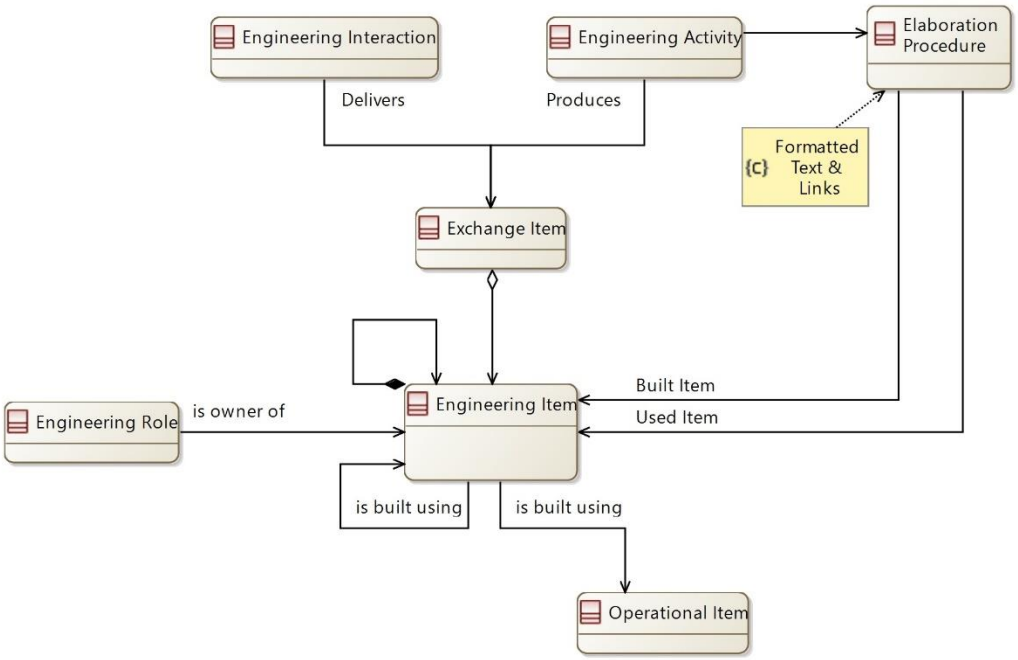

Figure 7: Meta-model of the "Engineering Activities for trustable AI" viewpoint

NOTE:

- On each "Engineering Interaction" between "Engineering Activities", one or more usages of data is/are described, as "Exchange Item" referencing used data.
- This allows specifying different states or characteristics of data according to their lifecycle and the places where they are used (to be described in "Exchange Item" and "Engineering Interaction").

C.3.5 Illustration

"AI-related Data Lifecycle" views are captured in the Operational Analysis (OA) perspective of Capella.

Mapping of the Viewpoint's concepts to Capella objects:

- "Engineering Interaction" maps to Capella OA's "Operational Interaction".
- "Engineering Activity" maps to Capella OA's "Operational Activity".
- "Elaboration Procedure" can be defined by using a context-aware description of an Operational Activity (including references to Engineering Items).
- "Exchange Item" maps to Capella's "Exchange Item".
- "Engineering Item" and "Operational Item" map to Capella's Classes involved as "Exchange Item Element" of an "Exchange Item".
- "Engineering Role" maps to Capella OA's "Operational Entity" or "Role".

The diagram below “Link between activities interactions and data” is only a first draft intended to illustrate a use of the concepts described in §C.3.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

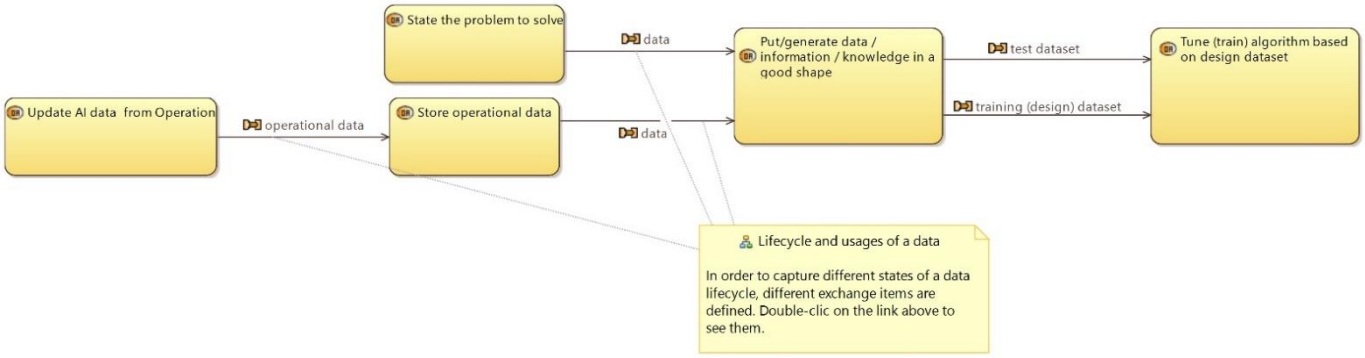

Figure 8: First draft of the view “Link between activities interactions and data”

C.4 Viewpoint “Risk on Trust due to Engineering”

C.4.1 Purpose & contents

- Identify major sources of bias or corruption brought by other engineering activities to inputs and outputs of AI engineering and data.

C.4.2 Justification

- Even if engineering of an AI Component is trustable per se, side effects of other engineering & development activities might bias or corrupt its definition.

C.4.3 Elaboration approach

- List engineering data & assets used and produced by AI engineering, contributing to Trust; identify risks of bias or corruption.
- List engineering activities dealing with former data & assets; identify related risks for AI.
- Define changes in activities and interactions to prevent from the identified risks on Trust.
- Include also development and production-related risks, similarly.

C.4.4 Concepts and meta-model

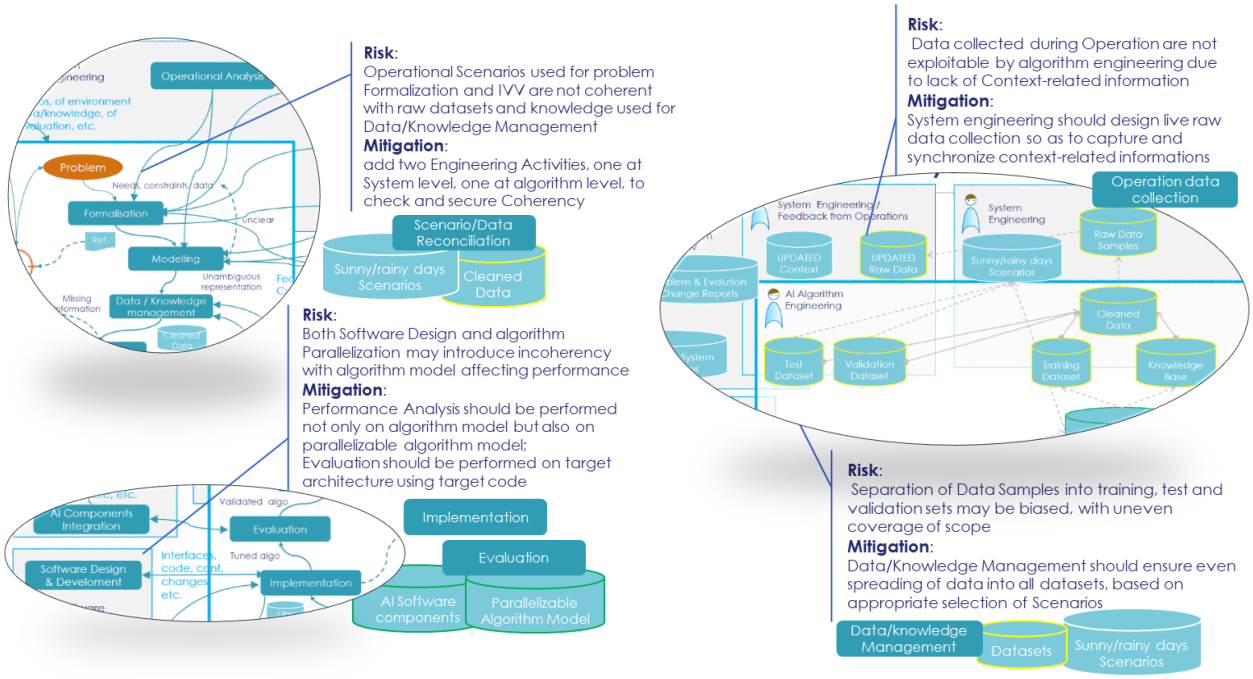

Figure 9: Concepts used in the "Risk on Trust due to Engineering" viewpoint

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

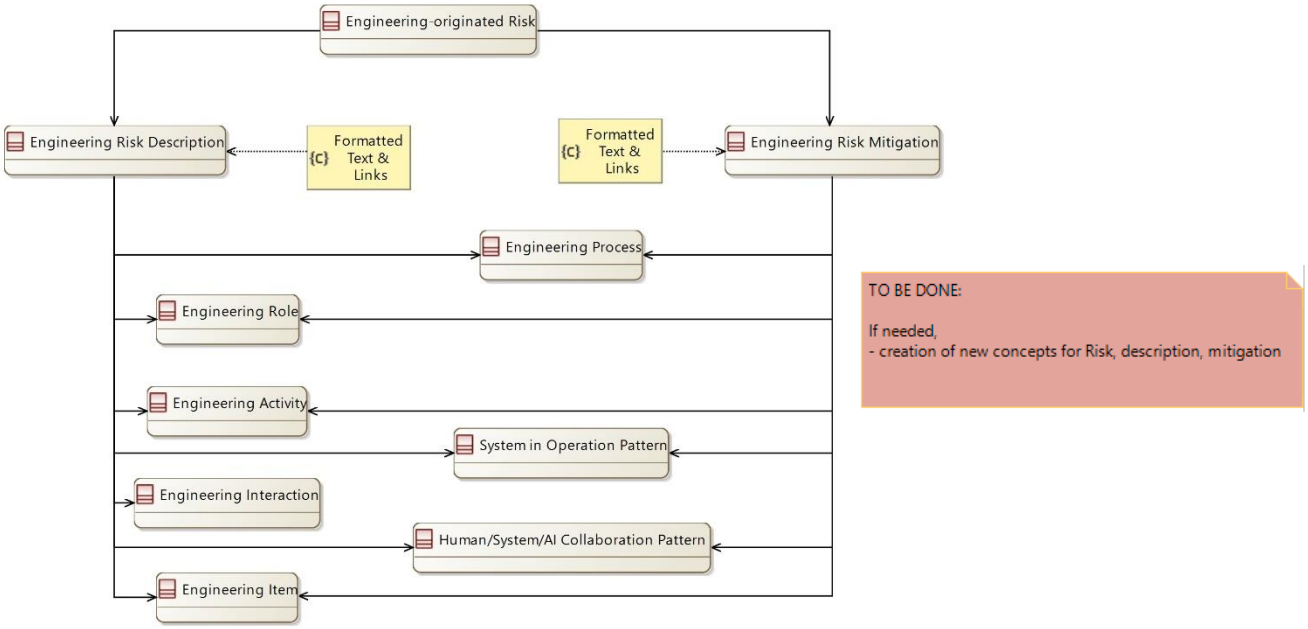

Figure 10: Meta-model of the "Risk on Trust due to Engineering" viewpoint

C.4.5 Illustration

“Risk on Trust due to Engineering” views are captured in any relevant perspective of Capella (depending on the objects on which risks apply).

Mapping of the Viewpoint’s concepts to Capella objects:

- “Engineering-originated Risk” maps on Capella’s “Constraint”.
- “Engineering Risk Description” maps on associated constraint description and links.
- “Engineering Risk Mitigation” maps on associated constraint description and links.
- “Engineering Process” maps to Capella OA’s “Operational Process”.
- “Engineering Role” maps to Capella OA’s “Operational Entity” or “Role”.
- “Engineering Activity” maps to Capella OA’s “Operational Activity”.
- “System in Operation Pattern” maps to Capella’s “Functional Chain”.
- “Engineering Interaction” maps to Capella OA’s “Operational Interaction”.
- “Human/System/AI Collaboration Pattern” maps to Capella’s Functional Chain.
- “Engineering Item” maps to Capella’s Class involved as “Exchange Item Element” of an “Exchange Item”.

The diagram below “Engineering Risk & Mitigation: Check and secure data coherency” is only a first draft intended to illustrate a use of the concepts described in §C.4.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

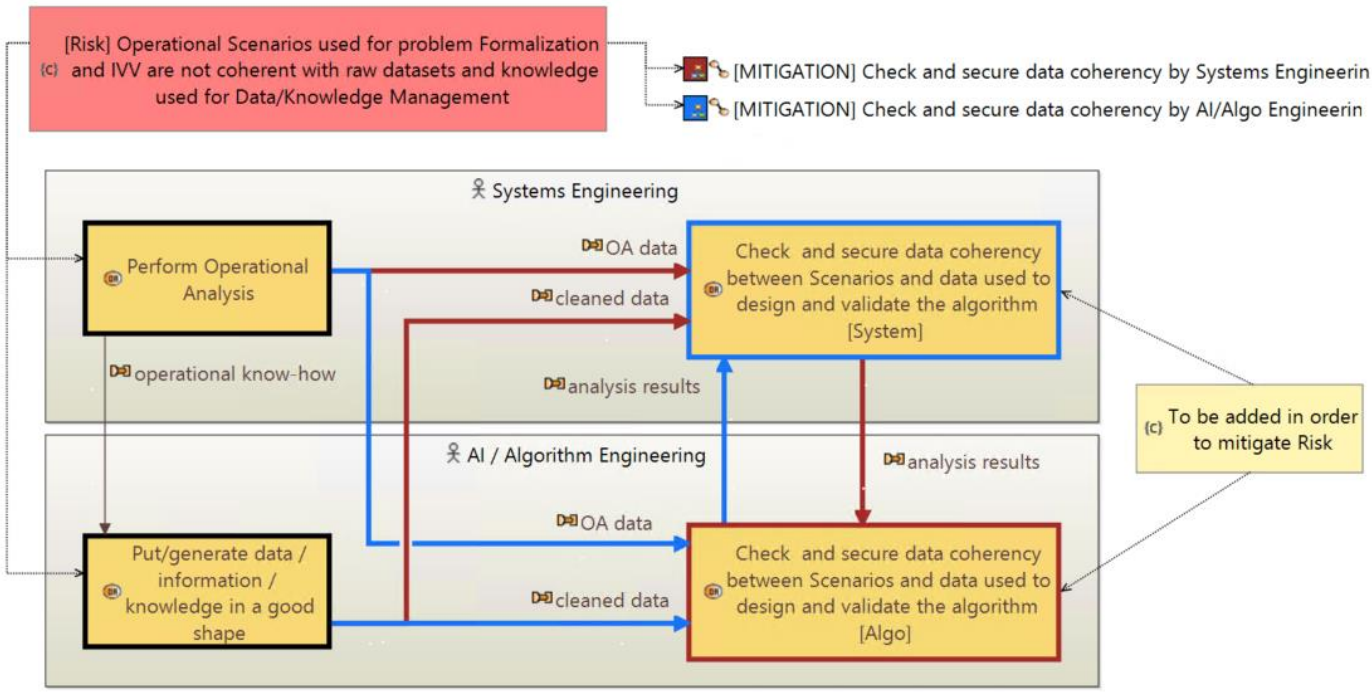

Figure 11: First draft of the view "Engineering Risk & Mitigation: Check and secure data coherency"

C.5 Viewpoint "Risk on Trust due to System in Operation"

C.5.1 Purpose & contents

- Identify major sources of bias or corruption brought by other system components interacting with AI components in Operation.

C.5.2 Justification

- Even if an AI Component is trustable per se, malfunctioning of other components might corrupt its inputs or misuse its outputs, or alter its behavior. (Consequences of AI Algorithms possible failures or malfunctioning are to be addressed also in "Trust on Performance, Safety, Security" Viewpoint.)

C.5.3 Elaboration approach

- List kinds of components interacting with AI components.
- Analyze possible failures or malfunctioning and their consequences on AI Algorithms and components.
- Define patterns (functional, architectural, technological) that could either prevent, detect of correct as needed.

C.5.4 Concepts and meta-model

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

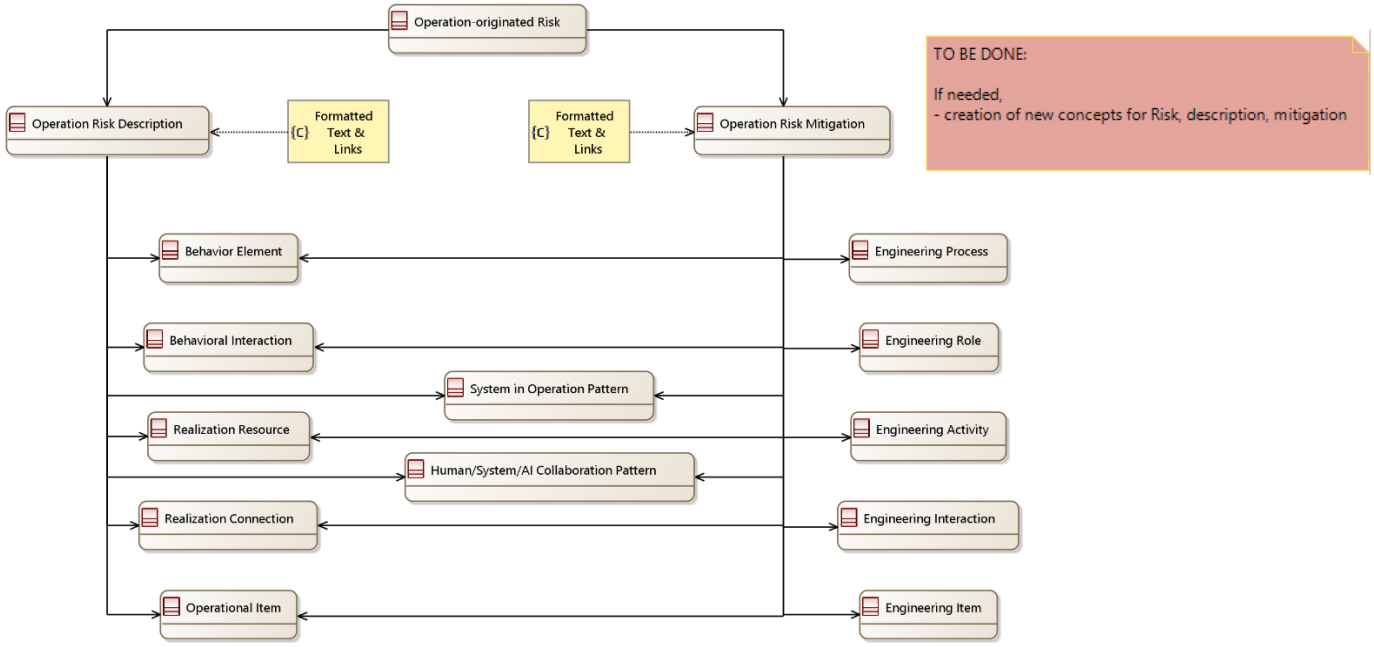

Figure 12: Meta-model of the "Risk on Trust due to System in Operation" viewpoint

C.5.5 Illustration

"Risk on Trust due to System in Operation" views are captured in any relevant perspective of Capella (depending on the objects on which risks apply).

Mapping of the Viewpoint's concepts to Capella objects:

- "Operation-originated Risk" maps on Capella's "Constraint".
- "Operation Risk Description" maps on associated constraint description and links.
- "Operation Risk Mitigation" maps on associated constraint description and links.
- "Behavior Element" maps to Capella LA's "Logical Function" or Capella PA's "Physical Function".
- "Engineering Process" maps to Capella OA's "Operational Process".
- "Behavior Interaction" maps Capella's "Functional Exchange".
- "Engineering Role" maps to Capella OA's "Operational Entity" or "Role".
- "System in Operation Pattern" maps to Capella's "Functional Chain".
- "Realization Resource" maps to Capella's "Behavioral or Node Implementation Component".
- "Engineering Activity" maps to Capella OA's "Operational Activity".
- "Human/System/AI Collaboration Pattern" maps to Capella's "Functional Chain".

- "Realization Connection" maps to Capella's "Component Exchange or Physical Link".
- "Engineering Interaction" maps to Capella OA's "Operational Interaction".
- "Engineering Item" and "Operational Item" map to Capella's Class involved as "Exchange Item Element" of an "Exchange Item".

The diagram below "Operation Risk & Mitigation: Protect against input freeze" is only a first draft intended to illustrate a use of the concepts described in §C.5.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

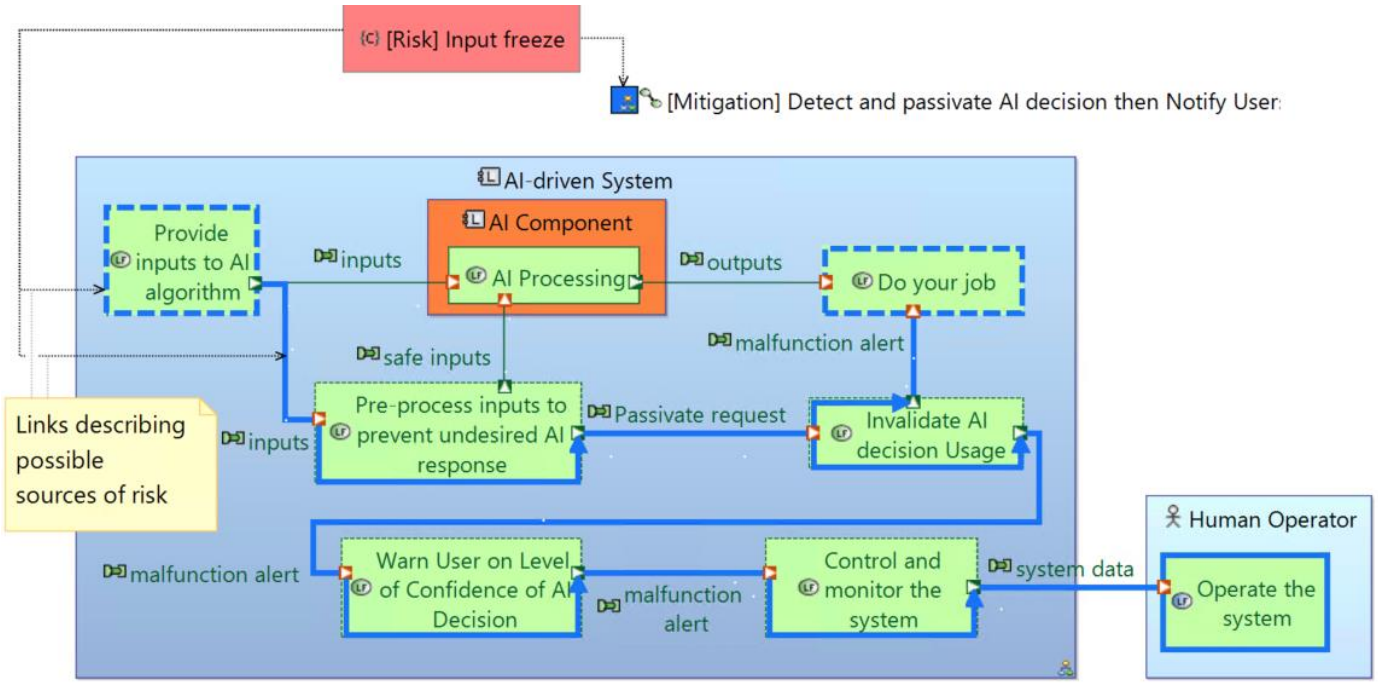

Figure 13: First draft of the view "Operation Risk & Mitigation: Protect against input freeze"

C.6 Viewpoint "Trust through System Behavior"

C.6.1 Purpose & contents

- Define major system capabilities needed to ensure Trust in Operation.

C.6.2 Justification

- Engineering per se cannot be sufficient to ensure trustable AI performance. Specific functional, architectural and technological capabilities are needed to secure the expected services in Operation.

C.6.3 Elaboration approach

- Based on former Viewpoints analysis, capitalize system functions, users actions and activities, architectural patterns etc., to secure behavior in operations.

C.6.4 Concepts and meta-model

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

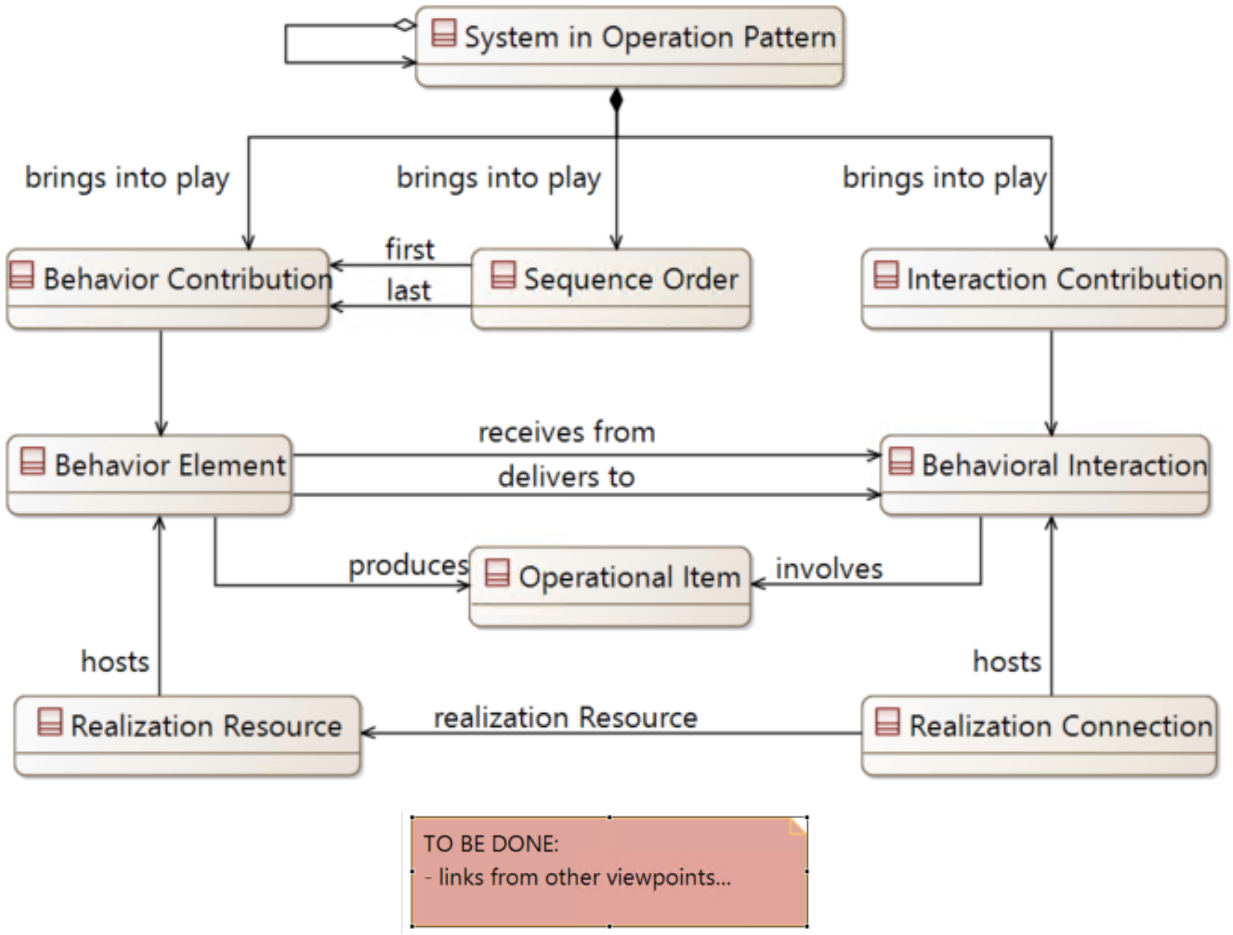

Figure 14: Meta-model of the "Trust through System Behavior" viewpoint

C.6.5 Illustration

"Trust through System Behavior" views are captured either in the Logical Architecture (LA) or in the Physical Architecture (PA) perspective of Capella.

Mapping of the Viewpoint's concepts to Capella objects:

- "System in Operation Pattern" maps to Capella's "Functional Chain".

- “Behavior Contribution” maps to Capella’s “Logical or Physical Function involvement”.
- “Sequence Order” maps to Capella’s “Sequence Link”.
- “Interaction Contribution” maps to Capella’s “Logical or Physical functional exchange involvement”.
- “Behavior Element” maps to Capella LA’s “Logical Function” or Capella PA’s “Physical Function”.
- “Behavior Interaction” maps Capella’s “Functional Exchange”.
- “Operational Item” maps to Capella’s Class involved as “Exchange Item Element” of an “Exchange Item”.
- “Realization Resource” maps to Capella’s “Behavioral or Node Implementation Component”.
- “Realization Connection” maps to Capella’s “Component Exchange or Physical Link”.

The diagram below “Operation risk & mitigation: Trust in case of platform failure” is only a first draft intended to illustrate a use of the concepts described in §C.6.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

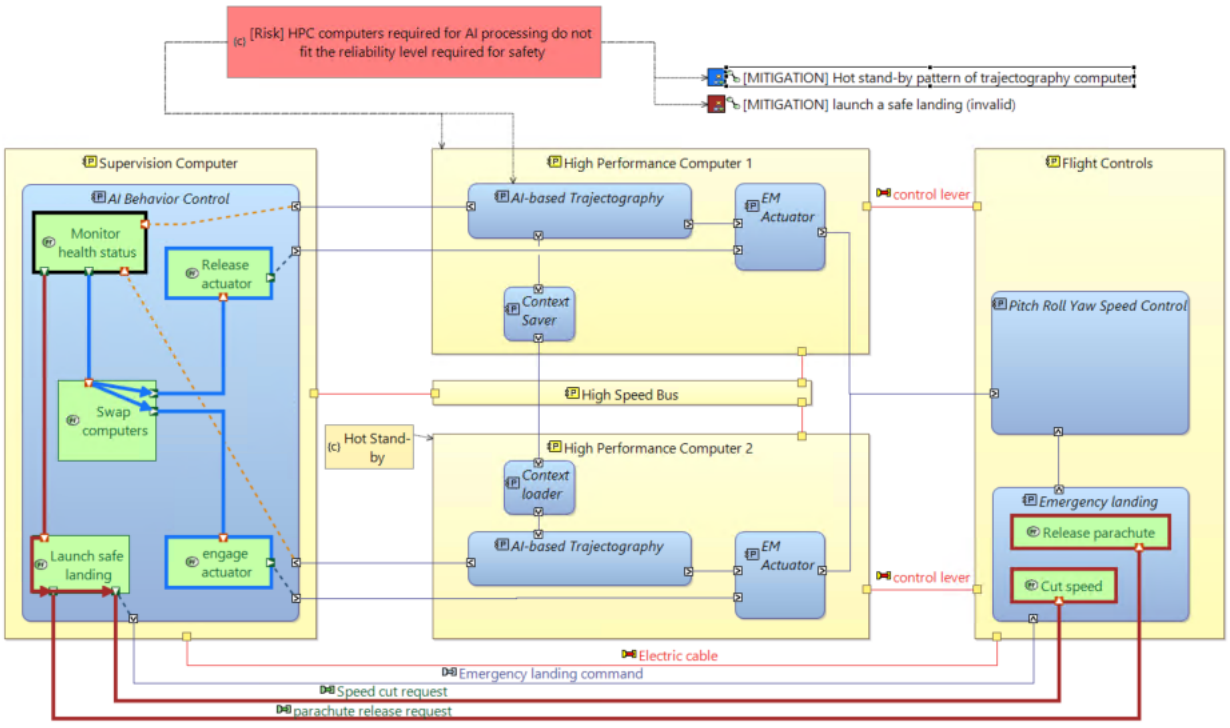

Figure 15: First draft of the view “Operation risk & mitigation: Trust in case of platform failure”

C.7 Viewpoint “Trust through Human/System/AI Collaboration”

C.7.1 Purpose & contents

- Define expectations of human stakeholders, their role and workshare with System & AI, in delivering the expected services in a trustable manner.

C.7.2 Justification

- Humans are a cornerstone to ensure Trust, and can also hinder or threaten Trust.

C.7.3 Elaboration approach

- Define major human stakeholders involved in operations, their possible contributions or obstacles to delivering the service.
- Define their activities, the information that they need or produce.
- Define corresponding AI and system related functions with which they interact, associated exchanges, non-functional properties, and typical interaction scenarios.

C.7.4 Concepts and meta-model

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

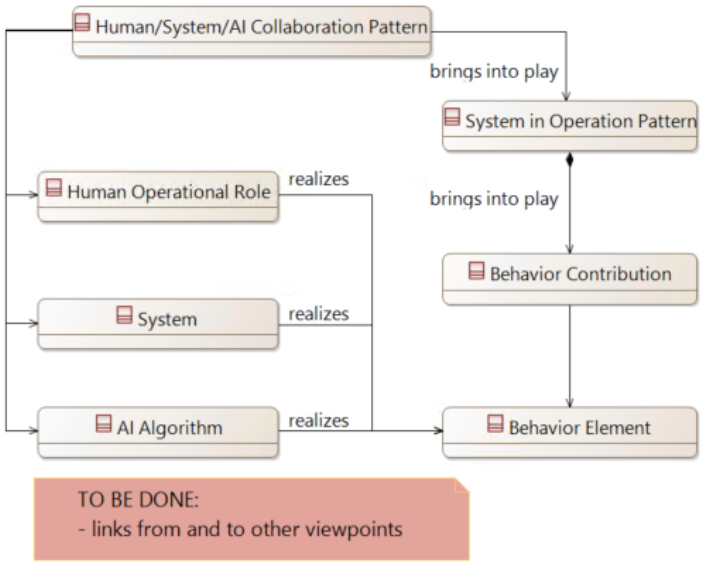

Figure 16: Meta-model of the “Trust through Human/System/AI Collaboration” viewpoint

C.7.5 Illustration

"Trust through Human/System/AI Collaboration" views are captured either in the Logical Architecture (LA) or in the Physical Architecture (PA) perspective of Capella

Mapping of the Viewpoint's concepts to Capella objects:

- "Human/System/AI Collaboration Pattern" maps to Capella's "Functional Chain".
- "System in Operation Pattern" maps to Capella's "Functional Chain".
- "Behavior Contribution" maps to Capella's "Logical or Physical Function involvement".
- "Behavior Element" maps to Capella LA's "Logical Function" or Capella PA's "Physical Function".
- "Human Operational Role" maps to Capella's "Actor".
- "System" maps to Capella's "System".
- "AI Algorithm" maps to Capella's "Logical or Physical Behavioral Component".

The diagram below "Human System AI collaboration pattern" is only a first draft intended to illustrate a use of the concepts described in §C.7.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

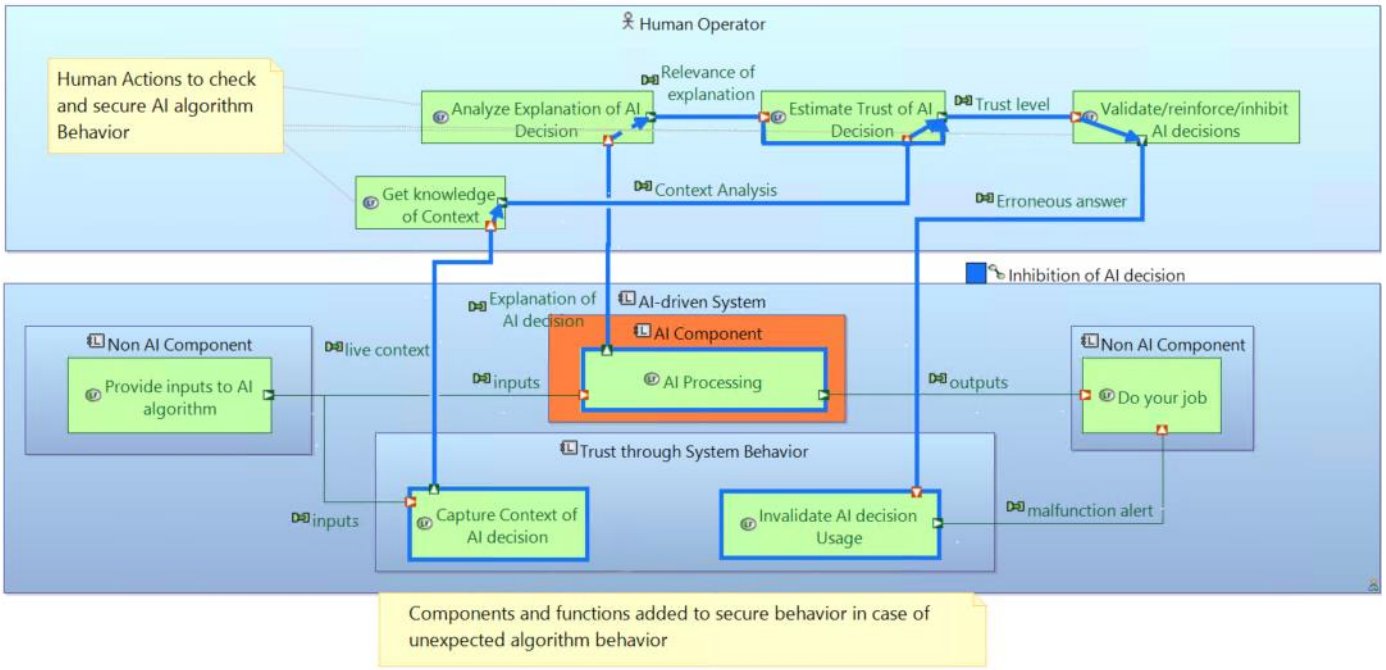

Figure 17: First draft of the view "Human System AI collaboration pattern"

C.8 Viewpoint “Integration of AI functions”

C.8.1 Purpose & contents

- Characterize and address specific concerns related to integrating one or more AI functions in system target context.
- Deliver guidance on the way to manage each concern.

C.8.2 Justification

- Prerequisites and consequences of AI components integration in their environment affect many engineering stakeholders and activities, and require taking specific precautions.
- The way all engineering teams cooperate in integrating AI functions is both a key enabler for success, and possibly a source of failure.

C.8.3 Elaboration approach

- Consider this viewpoint as transverse to others.
- Revisit the other viewpoints from the following perspectives:
  1) integration of data/knowledge and algorithm, up to target platform ;
  2) cooperation of several AI functions altogether ;
  3) integration of these functions in a distributed system architecture: prerequisites, consequences.
- Characterize and complement former viewpoints elements to provide guidance for achieving/securing this integration.

C.8.4 Concepts and meta-model

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

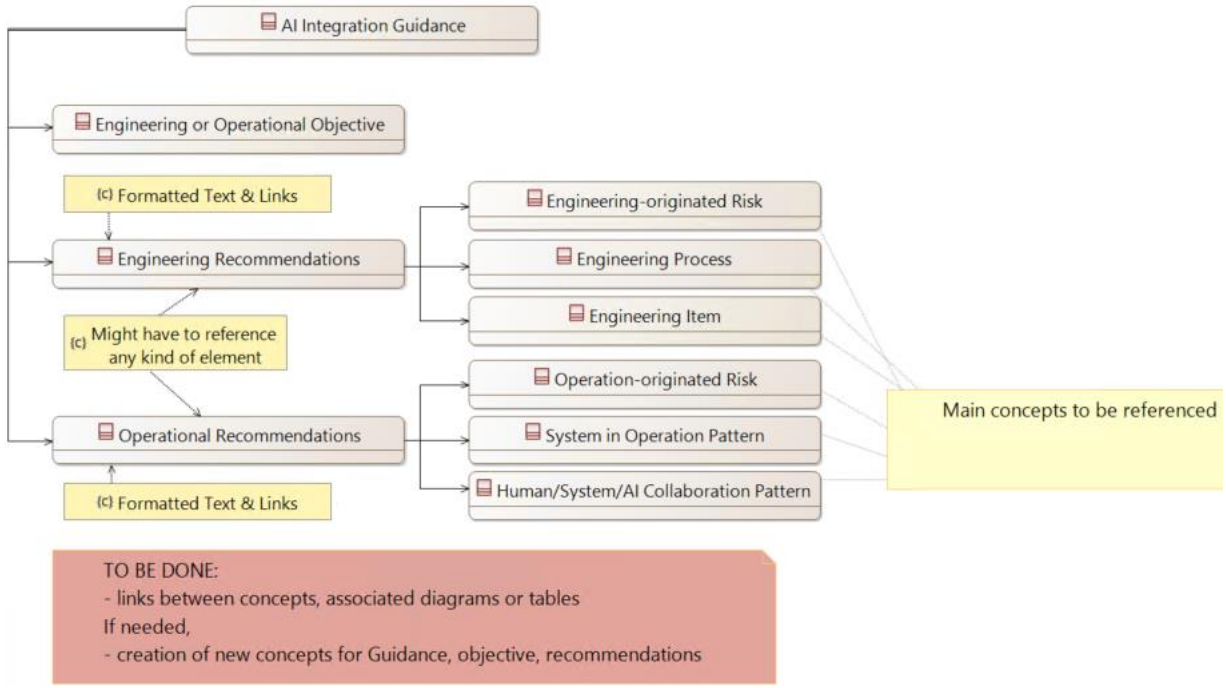

Figure 18: Meta-model of the "Trust through Human/System/AI Collaboration" viewpoint

C.8.5 Illustration

"Integration of AI functions" views are captured in the Operational Analysis (OA) perspective of Capella.

Mapping of the Viewpoint's concepts to Capella objects:

- "AI Integration Guidance" maps on Capella's "Operational Capability".
- "Engineering or Operational Objective" maps on associated constraint description and links.
- "Engineering and Operational Recommendations" map on Capella's "Operational Capability" and associated constraint description and links.
- The other concepts are mapped in previous viewpoints.

Such a view would be similar to the illustration in §C.9.5.

C.9 Viewpoint “Trust on Performance, Safety, Security”

C.9.1 Purpose & contents

- Define main needs, contributions and obstacles regarding Trust applied to AI decision performance, safety, and security of the global solution including AI.

C.9.2 Justification

- Safety, Security and probably AI Trust are likely to benefit from same kinds of engineering approaches.
- The way to address them in engineering is similarly transverse to other viewpoints.

C.9.3 Elaboration approach

- Consider this viewpoint as transverse to others.
- Analyze each of the other viewpoints from the three perspectives: performance, safety, security.
- Analyze kinds of threats and formalize possible solutions/mitigations of each viewpoint element; formalize its contribution to trust.
- Define global trust elaboration process(es) integrating all these contributions.

C.9.4 Concepts and meta-model

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

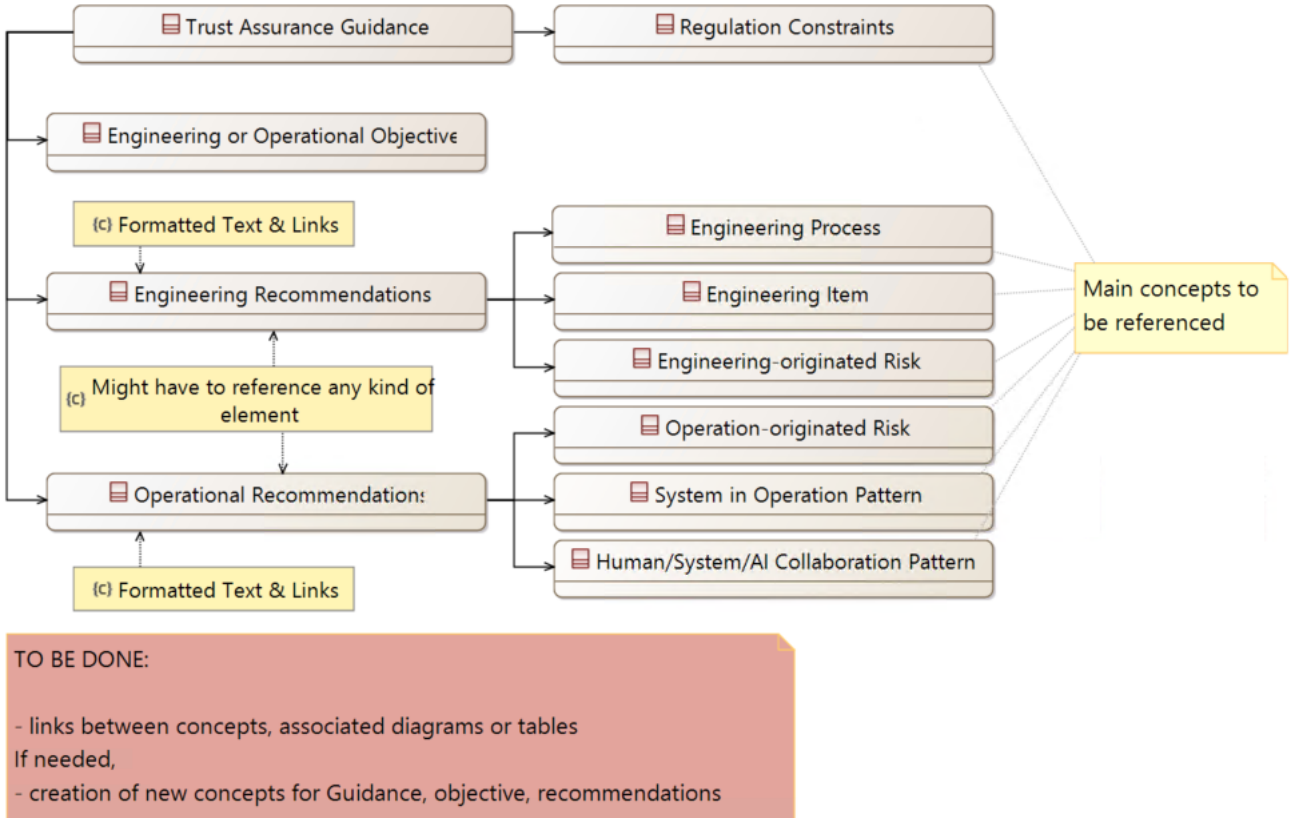

Figure 19: Meta-model of the "Trust on Performance, Safety, Security" viewpoint

C.9.5 Illustration

"Trust on Performance, Safety, Security" views are captured in the Operational Analysis (OA) perspective of Capella.

Mapping of the Viewpoint's concepts to Capella objects:

- "Trust Assurance Guidance" maps on Capella's "Operational Capability".
- "Engineering or Operational Objective" maps on associated constraint description and links.
- "Engineering and Operational Recommendations" map on Capella's Capability and associated constraint description and links.
- The other concepts are mapped in previous viewpoints.

The diagram below “System level securing of AI behavior” is only a first draft intended to illustrate a use of the concepts described in §C.9.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

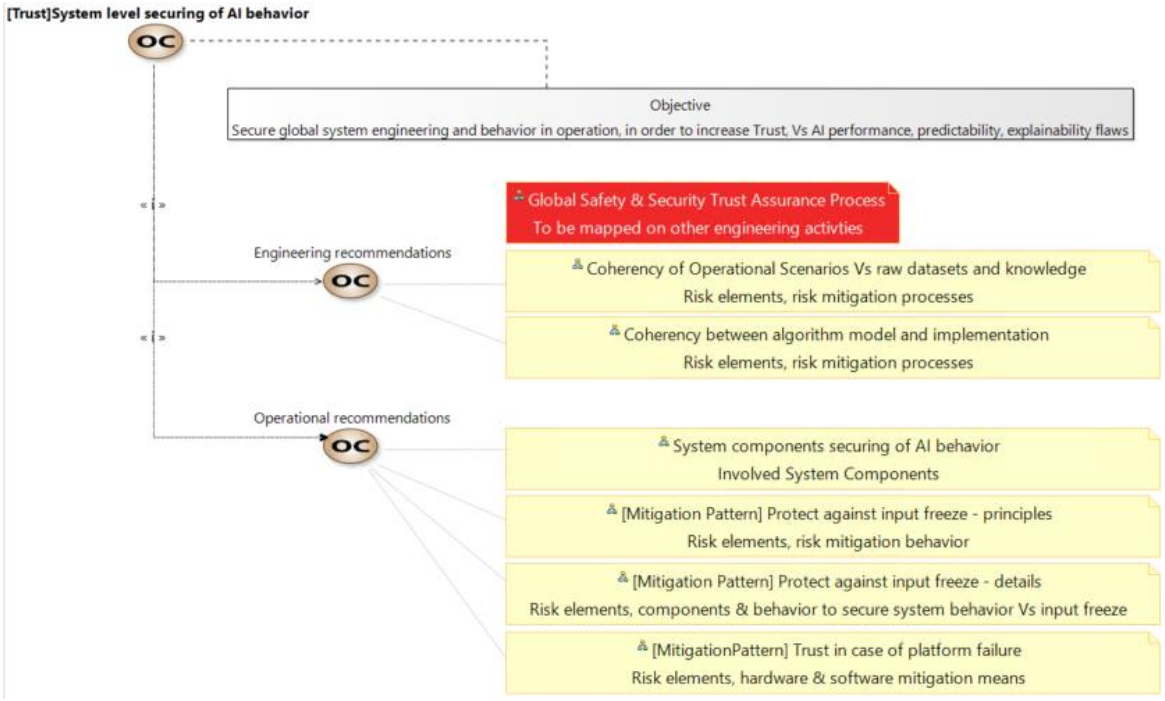

Figure 20: First draft of the view “System level securing of AI behavior”

C.10 Viewpoint “Engineering Lifecycle Management”

C.10.1 Purpose & contents

- Define main needs, contributions and obstacles regarding Trust applied to AI decision performance, safety, and security of the global solution including AI.

C.10.2 Justification

- Evolution of context and environment may lead to degrading performance and trust of AI and system, hence requiring refactoring.
- Engineering lifecycle has consequently to integrate continuous feedback, and to extend up to disposal.

C.10.3 Elaboration approach

- Consider this viewpoint as transverse to others.
- Analyze which system and engineering (including AI) inputs may make necessary to revisit the design and the development.

- Define processes and complementary engineering activities, required data, etc., to detect evolutions of context, analyze their consequences, and make design & development evolve accordingly.

C.10.4 Concepts and meta-model

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

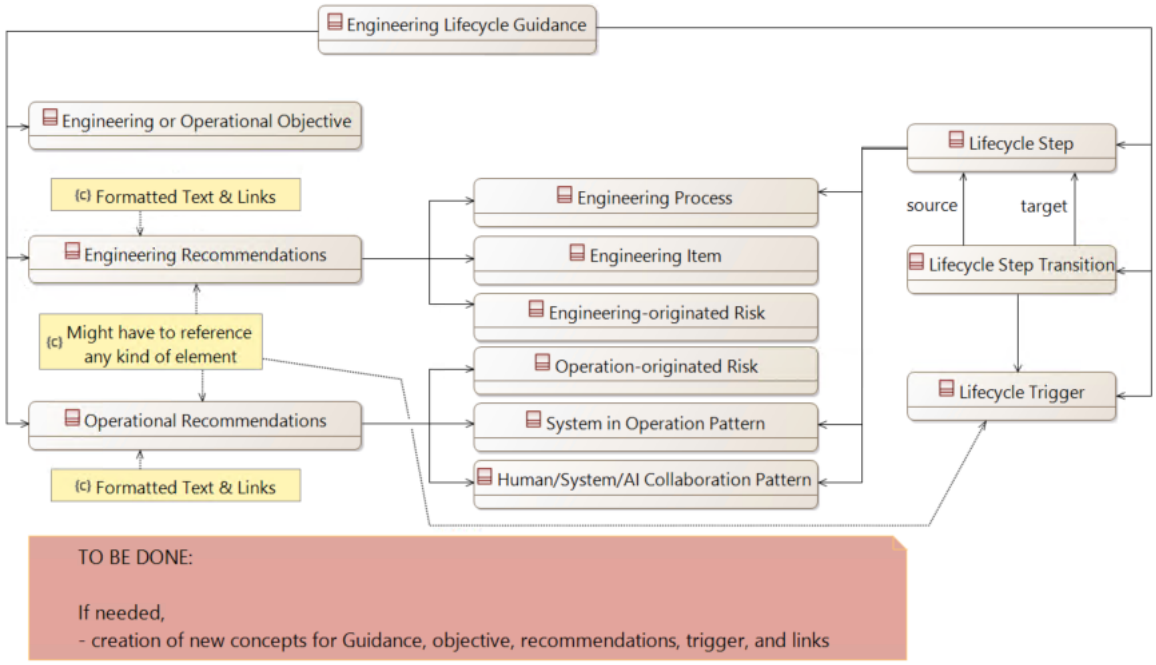

Figure 21: Meta-model of the "Engineering Lifecycle Management" viewpoint

C.10.5 Illustration

"Engineering Lifecycle Management" views are captured in the Operational Analysis (OA) perspective of Capella.

Mapping of the Viewpoint's concepts to Capella objects:

- "Engineering Lifecycle Guidance" maps on Capella's "Operational Capability".
- "Engineering or Operational Objective" maps on associated constraint description and links.
- "Engineering and Operational Recommendations" map on associated constraint description and links.
- "Lifecycle Step" maps on Capella's "Mode" or "State".
- "Lifecycle Step Transition" maps on Mode or State transition.
- "Lifecycle Trigger" can map to free text or any model element.

The diagram below “AI/Algorithm Engineering phases” is only a first draft intended to illustrate a use of the concepts described in §C.10.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

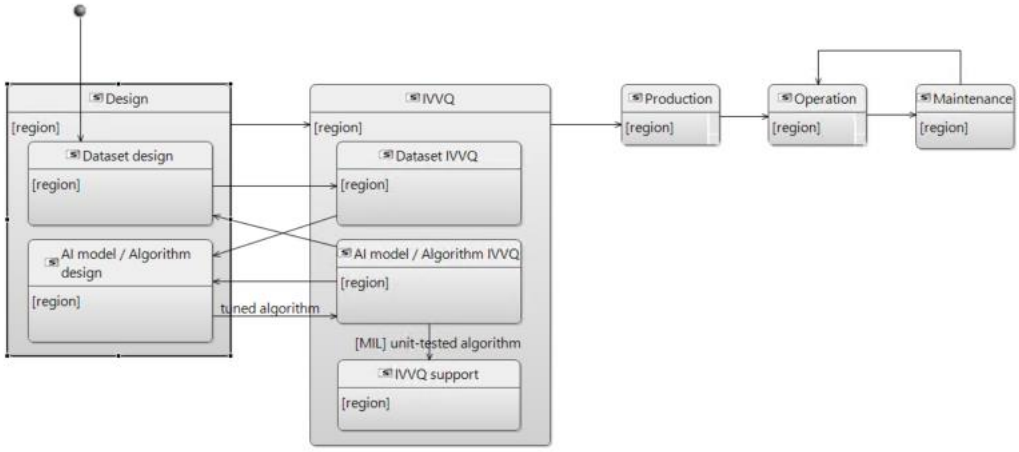

Figure 22: First draft of the view “AI/Algorithm Engineering phases”

C.11 Viewpoint “Engineering Environment Capabilities”

C.11.1 Purpose & contents

- Define the tooling support required to make trustable AI-based systems engineering feasible, scalable, efficient and secure.

C.11.2 Justification

- Tooled assistance, mechanization and automation to confront and reconcile all constraints of engineering stakeholders, activities, data, will be a key enabler in order to master complexity and scale.

C.11.3 Elaboration approach

- Consider this viewpoint as transverse to others.
- For each practice, activity, data and data transformation identified in former viewpoints, define user-level applicative services to support them.
- Define generic, transverse technical services likely to contribute to implementing former applicative services by smart assembly.

C.11.4 Concepts and meta-model

The description below (shown as a Class Diagram in Capella), is only a notional meta-model (set of related concepts); implementation meta-model could be based on existing standards when appropriate.

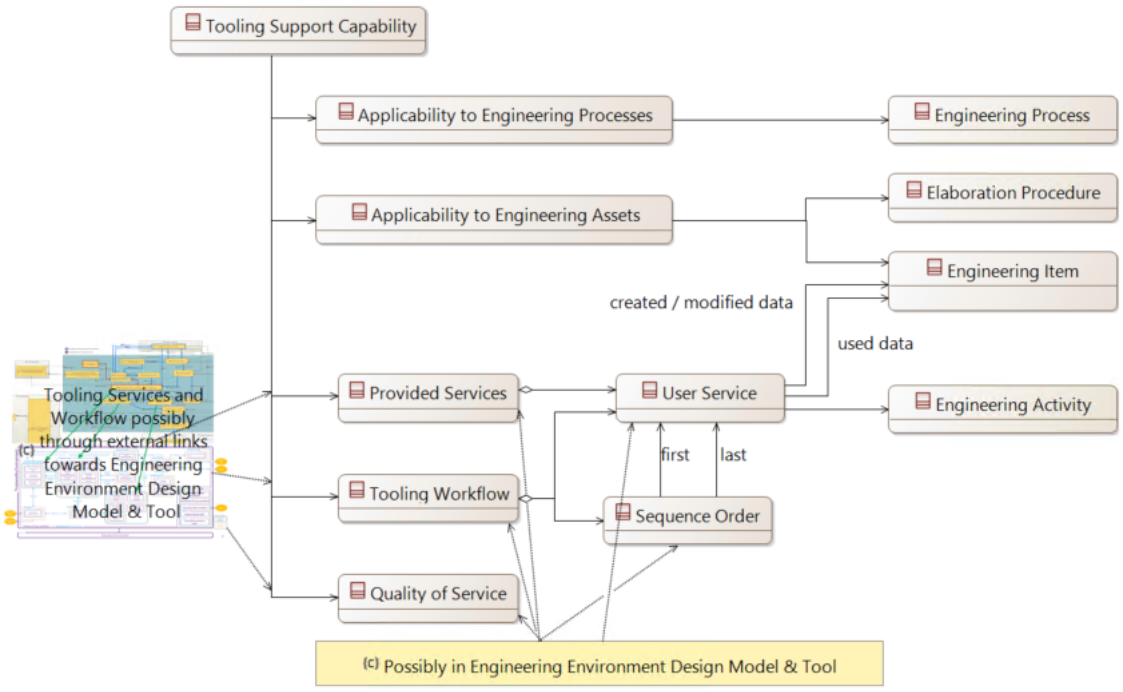

Figure 23: Meta-model of the "Engineering Environment Capabilities" viewpoint

C.11.5 Illustration

The diagram below "Link with tooling services and workflow" is only a first draft intended to illustrate a use of the concepts described in §C.11.4. It will be modified/corrected/improved during the remainder of the **Confiance.ai** project.

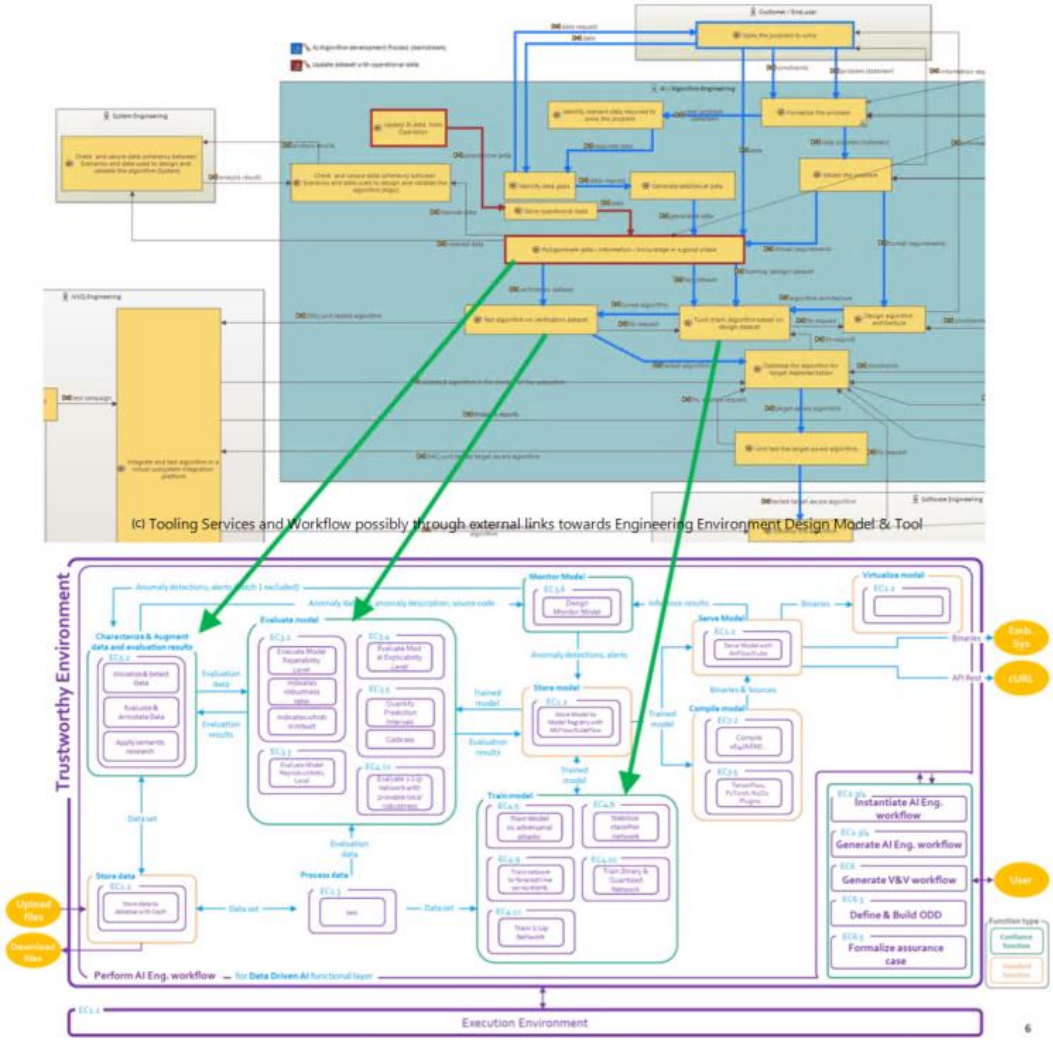

Figure 24: First draft of the view "Link with tooling services and workflow"

D. Conclusion

This document provides the status of **Confiance.ai** EC2 project on the Trustable AI Engineering Definition Framework. From now on, this Framework is going to capture outputs of other **Confiance.ai** projects, especially the ones applied to **Confiance.ai** use cases. For instance, we are already taking into account feedback from **Confiance.ai** EC6 project in order to integrate their Assurance Case approach into this Framework.

The way forward for the following months is to enrich and consolidate the current draft of viewpoints and views definitions, first in the scope of Batch 1 (Machine Learning on image and temporal series data for less critical application). Then, the scope will be extended to Batch 2 (Machine Learning AI and knowledge based AI on other type of data, such as text or audio, for medium critical application) and Batch 3 (hybrid AI critical application).

Once a consistent and consolidated capture is finished for a given Batch, this Trustable AI Engineering Definition Framework should serve as part of the specification for **Confiance.ai** EC1 project. Indeed, the Trustable AI Engineering Framework will have integrated the knowledge captured during its definition phase and should guide industrial end-users for the usage of the workbench (methods, strategies to apply) that will be delivered by **Confiance.ai**.

Our partners