Confiance ai

x.fr

Contributors

Name	Organisation
Bertrand Braunschweig Fabien Tschirhart	IRT SystemX
Sylvaine Picard	SAFRAN
EC1, EC2, EC6 end EC7 members	

EC1: Federative environment

Document control

Revision version	Modifications		
V4.0	Changes from Christophe	Draft	See above

Synthesis of Industrial Practices

Confiance.ai-REF

Contributors

(many names)

1. Executive Summary

This document is the main outcome of a survey sent to all industrial partners of Confiance.ai aimed at establishing their current use of tools and methodologies for trustworthy AI, and at identifying the needs they express on the subject. Each industrial partner received a set of 140 questions intended for one or more business profiles and corresponding to both a specific engineering phase and an object of interest. 52 completed questionnaires were obtained.

The conclusions are organized in eighteen sections covering general matters, specific objects of interest (e.g., the ODD, knowledge bases etc.), and a sequence of engineering phases. It is completed with an appendix provided by project EC6 with a specific analysis of IVV subjects, and an appendix with the list of questions.

The following is a summary of the eighteen sections, detailed in section 3 of this report.

Confidence

This part of the questionnaire deals with general considerations about trust/confidence, and about the services to be delivered by the trustworthy environment (also named "confidence environment" or "federative environment"). It emphasizes the need for the program Confiance.ai to communicate its analyses and achievements to its members, in order to share a common view of the terminology used, of the available technologies and methods (e.g., ALTAI is seldom used, and probably unknown to many), and of the major publications on the subject. It is interesting to note that most bibliographic references mentioned are norms and standards, even if there is abundant other literature on the subject. The main services expected by the participants are linked to certification and validation on representative use cases, with supporting methods and tools.

Generalities

This section deals with questions about the use of trustworthy AI components by the partners. It is shown that many partners already implemented AI components in a solution, up to operational use for a small number of them; moreover, the plan is to do it for more and more applications in almost all activities of the companies, to improve existing system or to provide new functionalities: e.g., perception, decision support, question answering, mission support, command and control systems etc. Foreseen applications are both batch and streaming/real time, with some optimism on the level of autonomy that can be reached in the next five years.

Operational Design Domain (ODD)

This is a difficult subject. At the moment, there is a lack of methods and tools for the definition of the ODD and for detecting when systems exit the ODD in operation. Some partners use domain-specific methodologies for the definition of the ODD, but we have no clues on what happens in operation.

Data engineering

For obvious reasons, data receives some attention from the respondents: many tools are mentioned for various aspects of the data lifecycle: feature engineering, data augmentation, training and validation, data engineering in general. Many off-the-shelf commercial tools are used as well as specific in-house tools developed by the partners often using python e.g., with scikit-learn. A number of data engineering methodologies are mentioned but one could say that in most cases the tools play the role of methodologies.

Machine Learning

The domain of Machine Learning is supported by numerous commercial and academic tools and other technologies (e.g., benchmarks, criteria, methods). Our industrial partners use these technologies for their ML developments, and sometimes add in-house products for the same purpose. In upstream phases, no methodology seems to be used, but once engineers move to practical implementation, they are empowered by the richness of the current supply. Tools such as Tensorflow, Keras, scikit-learn, Pytorch are often mentioned, as well as popular benchmarks and criteria. Bottom line, there does not seem to be any complaint about the dependencies over these products.

Knowledge-based approaches

Knowledge-based approaches are little used by industrial partners. We received only a few answers on the subject, and most often people do not know about the methods and tools in this domain. We see a tiny number of mentions of ontology tools such as Protégé, some in-house technology, and that's about it. The survey was done during Batch 1 of the program, which is devoted to data-based AI and machine learning solutions. It might be useful to do a second round in 2023 to see if things change significantly since Batches 2 and 3 will include KB use cases and solutions.

Human-Computer Interaction (HCI)

The subject of HCI is a very rich one. The respondents identify three modes of collaboration between AI systems and humans. In the three modes, the demands are for AI systems to be reliable, reactive, to reduce the worker's cognitive load, to be reproducible, transparent, in short to behave as a trustworthy co-worker. The major risk identified is when an AI system reproduces the same errors over and over again, which corresponds to the famous saying "trust take the stairs up and the elevator down". All these issues are to be taken seriously in the program.

Safety

The main safety concerns raised by industrials are the potential errors of an AI system and their consequences. Guaranteeing precision and generalization, linearity of inference, characterizing the uncertainty of answers are obvious needs. Several standards have been designated, depending on the industrial domain and reveal that industrials almost systematically work in standard controlled environment.

Interoperability

The Federative Environment has to be compatible with common infrastructure stacks such as Kubernetes and HPC. Means it must be deployable and runnable on both system (on a virtual machine or an HPC environment) therefore the AI toolchain (such as Keras, TensorFlow, PyTorch, … for machine-learning dedicated tools) must be compatible with the execution environment regardless of the target infrastructure. Confidentiality and cybersecurity are two main constraints that should not be neglected when solving interoperability needs. For instance, non-sovereign code should not be used.

Lifecycle management

Lifecycle management is an activity that seems to be little addressed in view of the few responses obtained. Sometimes in the course of investigation, the lifecycle approach is generally relegated to a single tool (MLFlow and Git are often used for the model) and sometimes ignored.

Requirements engineering

The requirements are expressed in terms of performance, robustness, explainability, interpretability. When dealing with embedded systems, the usual constraints are mentioned: size, power consumption, real time performance. The methods and tools used for requirements engineering of AI components are the ones currently used by industrials for traditional components.

Design

This section examines the use of formal methods, processes and methodologies and the supporting tools in the design of AI components and systems. There is little use of such approaches in current practice, and consequently there is an obvious need for Confiance.ai and other parties to deliver guidelines and tools for this purpose. When companies integrate components from third parties (ie. suppliers) they tend to ask for guarantees of some sort in the contracts.

Algorithmic Engineering

Too few details are provided to extract any information other than the non-use of algorithmic engineering in the AI process.

Assessment

In general, the assessment method of AI-based systems is partly based on the application case: while there are indeed criteria for evaluating individual AI components of a system, the integration of humans or the hybridization of the two proves less covered. There is definitely a lack of methodology for characterizing an AI system, which currently seems to be done on a case-by-case basis, sometimes in line with a certification process, but very often in an empirical way with a clear confusion between system performance and trustworthiness. As it stands, the decision not to integrate AI into a solution is based on the fact that the IVVQ or criticality criteria of the function (with respect to the DAL) underlies a risk that the added value cannot compensate.

Deployment

This section mainly addresses deployment for embedded systems, since deployment on large-scale infrastructures or cloud do not seem to be a subject of concern. Most applications seem to be standalone on embedded hardware without connection to a "home-base" cloud. Therefore, edge computing is not mentioned as a need for the moment, although some might think that it will be the case in the near future. In terms of hardware targets, there is a great diversity from CPUs to GPUs including manycore, ASICs etc. Engineers use the tools and methods supplied by hardware manufacturers to port their AI components to the embedded platforms, but there are several associated weaknesses in terms of optimization, resource usage, energy consumption and performance estimation, also because the AI development platforms used do not provide such services. There is a lot of room for improvement on the deployment phase, which is one of the focuses of EC7. This project highlights the choice of the N2D2 tools which supports implementations on several target platforms.

Supervision/Monitoring

Generally speaking, the process of integrating the monitoring of AI-based systems does not seem mature. Various open-source tools are frequently mentioned (MLFlow, Kubernetes, Tensorboard, Prometheus, Grafana) as well as tools delivered with AI platforms (Azure, AWS Sagemaker), sometimes in-house tools monitoring specific system indicators are mentioned.It is quite clear that no formal monitoring methodology is currently applied.

Certification

It seems clear to everyone that certification of AI component will have a major impact on the current practices, for methodologies as well as for tools used in the development chain, essentially because there is little or no use of certification tools and methods for AI systems at the moment. Engineers are keen to apply such changes if the added value is demonstrated. However they ask to have some freedom of choice, not to be constrained by the requirements of certification tools and methods. Assurance cases are not known or considered not applicable.

Other questions

This section gathers questions that address complementary aspects of the program and do not fit in the previous sections.

- which engineering phase raises more challenges for AI? The answer is "all'.
- what criteria for adopting AI solutions: more or less all criteria one can think of.
- how to assess statistical performance: same as for conventional components
- quality assessment for annotations: double labelling with sometimes a tool to help
- adversarial attacks and forensics: no solution used as of now.

2. Introduction and main statistics

This document summarizes the results of the analysis work carried out as part of the inventory of the engineering tools and methodologies used by the industrial partners of Confiance.AI. Each partner received a set of questions; each of these questions was intended for one or more preestablished business profiles and corresponding to both a specific engineering phase and an object of interest.

Each industrial nominated a proxy in charge of the interface between the Confiance.AI collaborators conducting the study and the internal technical teams whose answers where required to conduct the analysis.

2.1. Engineering profiles

As mentioned earlier, each member of the technical teams who responded to the questionnaires was asked to indicate their correspondence or proximity to a particular engineering profile of reference. Below is the list of suggested engineering profiles.

Data scientist: covers the roles of modeler (selection, parameterization, fine tuning of models) for the implementation of a data driven approach to respond to a business problem.

AI algorithmic Engineer: covers the roles of AI algorithmic engineer, design/selection of algorithms and development or redevelopment (refactoring) of specific algorithmic bricks.

Safety Engineer: covers the roles, at the legal level, of person in charge of the safety of the designed system, of obtaining certification, ...

System Engineer: covers the roles of specifying and designing the system, driving the industrialization of the AI components in the system, ...

System IV&V Manager: covers the roles of system quality manager, driving the implementation of functionality/performance measurement component as well as procedures.

Knowledge Engineer: covers the role of capturing business rules, collecting them, qualifying them and updating them.

Data Manager: covers the roles of data administrator (archiving procedures, collecting, handling quality controlling, use) and responsible for their availability.

Human Cognition Engineer: covers the role of analysis, design and monitoring of human-machine interaction.

Software Engineer: covers the role of developer, integrator of the algorithm/model/AI brick in the system (industrialization), deployment in production of the system, ...

Engineering Workbench leader: covers the role of engineering workbench leader (in the broad sense of the word (system, SW, equipment, service)

2.2. General statistics

Out of a total of nine industrial partners consulted, seven returned at least one completed questionnaire and a total of **52 completed questionnaires were returned**, some with responses from several engineering profiles. From these 52 returned questionnaires, a total of **3583 responses to individual questions** were extracted. The following graph highlights the global distribution of answers according to the respondent's engineering profile.

The following graph allows us to identify the distribution of participation as a function of the job profile and thus of the biases, orientations and possible gaps in the study.

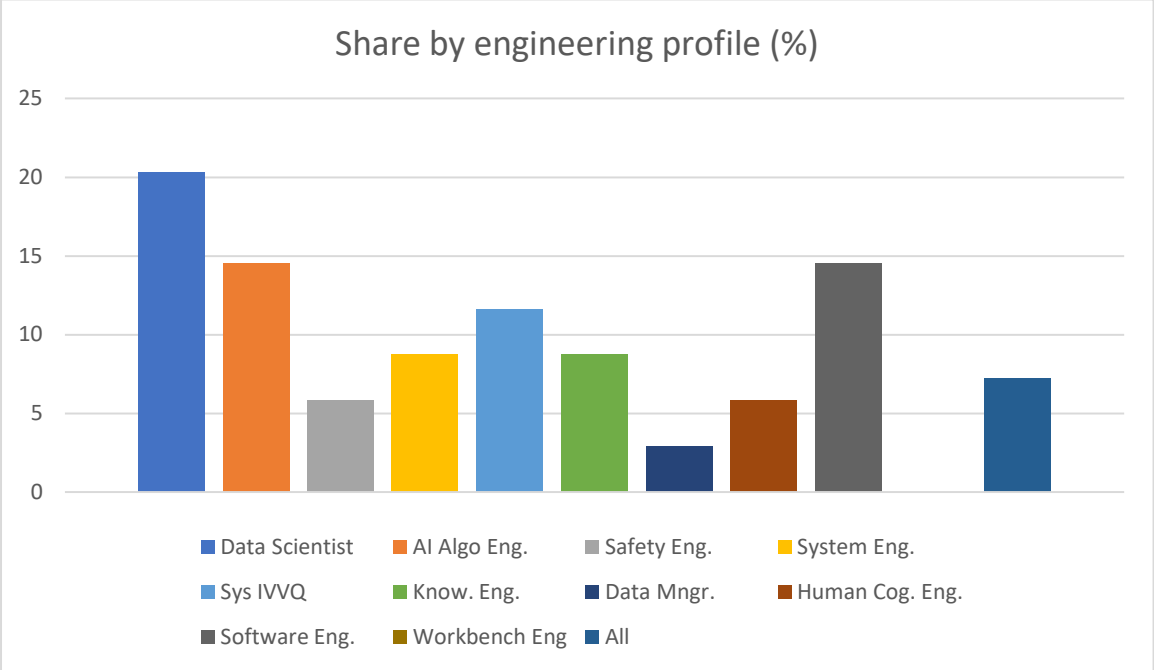

Fig 1. Global distribution of answers according to the respondent's engineering profile

Note: Answers whit no specific profile identified or mixed profiles indicated are tagged 'All' in the graph above.

For some questionnaires it happened that questions were left without any contribution. Through all the returned contributions, all the questions benefit from at least one answer by one of the industrial partners. This leaves no question unanswered.

In order to facilitate the processing of responses, each question was associated with one of 18 classes. The distribution of these classes, and therefore of the nature of the questions, was not homogeneous. The following table indicates for each class how many questions were associated with it. Each class is either an object of interest(oi), an engineering phase (ep), or other (oth).

Question classes	Number of questions
Algorithm (oi)	3
Assessment (ep)	6
Certification (ep)	8
Confidence (oth)	5
Data (oi)	16
Deploy (ep)	20
Design (ep)	6
Human Cognition Interaction (oi)	7
Interoperability (oth)	2

Knowledge-based (oi)	14
Lifecycle (oth)	2
Machine-learning (oi)	9
Operational Design Domain (oi)	6
Requirements (ep)	9
Safety (oth)	2
Supervision (ep)	6
General (oth)	12
Other (oth)	8

Table 1. Number of questions for each class

3. Synthesis

This section, the main section of the report, is organized as follows: we start (3.1 and 3.2) by the questions relative to the definitions of trust and confidence, and general questions about the current and future use of AI solutions by the partners. Then, we look at a number of objects of interest, namely the ODD, data, machine learning solutions, knowledge-based solutions (3.3 to 3.6), and some transvers aspects: human-computer interaction, safety and software interoperability (3.7 to 3.9). Then, all engineering phases are examined, starting by lifecycle management (3.10), followed by seven phases, namely requirements, design, algorithmic engineering, assessment, deployment, supervision/monitoring and certification (3.11 to 3.17). The last section 3.18 contains a few additional items not included in the previous ones.

3.1. Confidence

Summary

This part of the questionnaire deals with general considerations about trust/confidence, and about the services to be delivered by the trustworthy environment (also named "confidence environment" or "federative environment"). It emphasizes the need for the program Confiance.ai to communicate its analyses and achievements to its members, in order to share a common view of the terminology used, of the available technologies and methods (e.g., ALTAI is seldom used, and probably unknown to many), and of the major publications on the subject. It is interesting to note that most bibliographic references mentioned are norms and standards, even if there is abundant other literature on the subject. The main services expected by the participants are linked to certification and validation on representative use cases, with supporting methods and tools.

Details

127 Confidence, robustness, fairness, transparency, explainability, are terms often used in the context of AI. What definitions do you give to these terms in your organization?

Some partners give definitions to confidence, robustness fairness etc like:
Robustness: ability to resist to natural, adversarial perturbations, including domain drift
Confidence: self-evaluation of the quality of the prediction returned by the AI model, or capability to provide an adequate answer.
Explainability: what are the input elements that influence the most a prediction returned by the AI model.
These definitions are not always complete nor accurate. Other people cite norms or ask for more precisions to answer. This emphasizes the need of a shared nomenclature.

126 According to you, what are the main services expected from the "Confidence Environment"?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers.

The main services expected from the "Confidence Environment" are: - accredited engineering and development environment for certification - get some proof & evidences for models that are: easy to understand, fast to exploit, reliable in their results - robustness monitoring, characterization and adaptation Interoperable process representation - methods and supporting tools (as a shopping list) - validation on representative use cases that can be associated with internal ones - means to prepare the rule makers in the certification approach.

138 According to you, what should be the major components of the "Confidence Environment"? Methods and other guidance material (evaluation reports, processes…)?, Tools? Libraries? Frameworks? Benches? Software components. Other?

Methodology defining process steps with guidelines and associated tools; Tools for evaluation and confidence metrics; test tools, Processes for data management and segregation and validation tools; Traceability for AI models, decision making tractability, explainability (post-hoc or not) MLOps tools, robustness monitoring and adaptation platform. One interesting remark is that the proposed V&V tools and methodology should be recognized ones.

135 Are you using some assessment list for AI such ALTAI (Assessment List for Trustworthiness AI)? If yes, which requirement of the ALTAI represents the main challenge to you?

Generally not, except for one partner who uses it for three internal checklists: one focusing on data, one focusing on Training, one focusing on embeddability

136 What do you consider to be the best bibliographic references for trustworthiness in AI based systems?

The publication considered as best bibliographic references for trustworthiness in AI based systems are: - DIN SPEC 92001, - SOTIF, - ISO trustworthiness AI specification, - DEEL Certification Paper, - EUROCAE / EASA Guidelines

3.2. Generalities

Summary

This section deals with questions about the use of trustworthy AI components by the partners. It is shown that many partners already implemented AI components in a solution, up to operational use for a small number of them; moreover, the plan is to do it for more and more applications in almost all activities of the companies, to improve existing system or to provide new functionalities: e.g., perception, decision support, question answering, mission support, command and control systems etc. Foreseen applications are both batch and streaming/real time, with some optimism on the level of autonomy that can be reached in the next five years.

Detail

1 Have you already integrated an AI component in a solution?

Several solutions involving AI have been developed by the program partners. All of the methods mentioned by the respondents are approaches based on statistical learning. This seemed to be the best approach, explaining this choice. These solutions include visual perception applications (automatic parking, person re-identification, anomaly detection, situation awareness or runway exit prevention. Some of the solutions are implemented on cloud platforms, others are embedded in real-time hardware. In the latter case, the adequacy between algorithm and hardware has led to compromises.

2 What types of AI-based functions are you implementing od do you with to implement? (Command and control, Monitoring, control)

The main functions performed by the respondents are: automatic target recognition in aerial imagery, object detection and tracking, autoML, anomaly detection in images, semantic segmentation of images or information search. Those considered as promising are: ontologies for rule discovery in datasets.

60 In your business, what are the main classes of systems or services for which you expect AI to bring major benefits?

For "non-Thales" interviews: many classes of applications from pattern recognition, perception and scene understanding to decision support, recommendations & advisory systems, document classification etc. including critical and noncritical systems
For Thales interviews: just about everything. Image processing and vision. Time series. NLP. Command and control systems. Mission support. Autonomous systems. Etc.

61 In your business, what are the main classes of systems for which you actually plan to integrate AI?

For "non-Thales" interviews: same as Q60
For Thales interviews: Same as Q60 + trajectory planning, adaptive control, IT applications

122 For what reasons are you considering the usage of AI? To improve existing (non-ML) solutions? To provide new services? For other "strategical reasons"?

The usage of AI components is mainly considered by industrial partners to improve existing (non-AI) solutions and provide additional value when the non-IA solutions shows some limitations, and to provide new services or solution that cannot be achievable through non-AI systems.

62 What kind of computation mode would your AI-enabled solution perform? Streaming? Batch? Real-time?

Both aspects are addressed equally: there are batch (pre and post) as well as real time usages. Streaming and RT are not distinguished by the respondents.

139 If your AI system is aimed at providing enhanced autonomy, what level do you consider achievable within the next 5 years?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers.

Surprisingly, Algo Engineers did not respond to this question. Partners identify level 4 enhanced autonomy (for automobile) and L1A/B (for Aeronautics) as achievable within 5 years.

115 Are you already using AI techniques in your engineering process? Methodology

AI techniques are not used in today engineering process, except in very few cases (hyperparameter optimization, meta-learning) or still under validation (design assistance, knowledge retrieval)

116 Are you already using AI techniques in your engineering process? Tools

Same answer as Q115

63 How do you ensure the continuity between development/engineering and maintenance in operational conditions with presumably cyber-security constraints?

no answers were given on this question

120 According to you, on which part of process (development, verification and validation, qualification and certification) should we focus the effort (from the confiance point of view)?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers.

Only one answer. It seems that the notion of "confiance point of view" has not been understood.

137 Are you contributing to some workgroup on AI? What is your role in this workgroup?

Algo engineers are relatively active in the AI trustworthiness field. They participate in internal or external (e.g., ISO) working groups. It is not the case for all engineers' groups.

3.3. Operational Design Domain

Summary

This is a difficult subject. At the moment, there is a lack of methods and tools for the definition of the ODD and for detecting when systems exit the ODD in operation. Some partners use domain-specific methodologies for the definition of the ODD, but we have no clues on what happens in operation.

Detail

5 Do you use a particular methodology to characterize/design the operational domain?

When existing, the methodologies (and tools) for Operational Domain characterization or design seems specific to each domain: Some partners use Operational Analysis (with Arcadia tool), some translate operational scenarii directly into test procedures. No general methodology seems to emerge for the building of the Operational Domain.

6 What tools do you use to characterize/design the operational domain?

Some partners identify a lack of tools and methods to enable the validation (e.g completeness & correctness) of the operational area.

A lot of answers are that engineers rely on non-formal methods (pen and paper, interviews, habits of the team, experts…) and custom tools to annotate the data base and to visualize the results. Two methodologies are mentioned, the MBSE methodology and tools Capella (in particular at Thales), Enterprise Architect, CORE, or Cameo, and the Airbus method MOFLT (Mission/ Opération/ Fonction/ Logique/ Technique). Other tools have been listed once: Microsoft Office tools (Excel, Mindmap, Visio) and tWiki for documentation.

104 How do you ensure that your AI system still operates in its Operational Design Domain?

No information was provided by the respondents.

112 Do you use a specific methodology to monitor/track/supervise your operational domain/area?

Answers to this question, if any, have not been analyzed yet.

113 Do you rely on specific engineering tools to monitor/track/supervise your operational domain/area?

No dedicated tools were mentioned by the respondents.

114 Do you rely on specific solutions to monitor/track/supervise your operational domain/area?

Answers to this question, if any, have not been analyzed yet.

3.4. Data

Summary

For obvious reasons, data receives some attention from the respondents: many tools are mentioned for various aspects of the data lifecycle: feature engineering, data augmentation, training and validation, data engineering in general. Many off-the-shelf commercial tools are used as well as specific in-house tools developed by the partners often using python e.g., with scikit-learn. A number of data engineering methodologies are mentioned but one could say that in most cases the tools play the role of methodologies.

The detail is organized as follows: I) what data? ii) methodological questions; iii) questions about tools.

Detail

15 What methods and tools are you using to manage data, perform feature engineering (e.g., dataiku)?

No specific methods are described to manage data and perform feature engineering but a lot of tools have been mentioned. To perform features engineering various tools are mentioned from simple "handcraft" to the most automated: jupyter notebooks + classic libraries (pandas/sklearn/...), tensorboard, MATICS (LNE), AIF360, Matlab and Tale of data. However, they may not offer storage capabilities. To manage and store data some solutions are mentioned: IDACAP, Google Cloud Storage, but they may not offer feature engineering capabilities. Finally, to combine both storage, management and feature engineering capabilities 3 platforms are listed: "skywise" and "airbus cognitive services" from Airbus, AWS from Amazon, Dataiku. In-house tools are mentioned but without a lot of explanation about their functionalities.

35 Are you already using data augmentation techniques?

For computer vision use cases, classical data augmentation techniques are first used: geometric transformation, color transformations, noise, blurring. Some more advanced techniques are also used (e.g., GAN). The tools used are open-source libraries/modules such as: Albumentation lib, native functions of DL framework, custom augmentation in

python. Transformations functions can be defined with experts for better representativeness, but the representativeness is often not validated. For NLP use cases, data augmentation is difficult.

22 Would you consider using data coming from external sources in your ML development process?

Algo Engineers do not hesitate to use pre-trained models on large public datasets. They see their strong improvement of performance and do not answer to the trustworthiness aspects of these data. Data Engineers do not use external data.

14 Have you already considered using data ("examples") are part of a specification?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers.

Only in the automotive domain (automatic parking)

16 Do you use a particular methodology to specify the requirements to define the training and validation datasets?

At this time, Algo Engineers do not use any particular methodology to specify requirements on datasets. At best, they declare to use field expert knowledge combined with ML methods and principles (like Standford ML principles) even if some declare not to be concerned by this question as an Algo Engineer. It seems to be some confusion here. Concerning Data Engineers, we encounter among the answers the classical mistake that small test set is the norm (and what about statistical representativity of the measured performances?) Worse, some Data engineers make a confusion between supervised learning and data quality. Unsurprisingly Safety Engineers did not answer this question. They could/should have.

30 Do you use a particular methodology to build the training and validation datasets?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers.

No specific methodology detailed by industrial partners for the building of training and validation datasets.

32 Do you use particular methodology to integrate training and validation datasets?

Mentions are: ML methodologies, Stanford ML practices, Cross-validation.

134 Have you already implemented data engineering practices? Methods and processes. Methodologies.

Basic practices: Versioning, Atos checklist, CI/CD practices.

133 Have you already implemented data engineering practices? Tools

Basic tools: DVC, MongoDB, MLFlow; Atos FMLE's tools; CI/CD tools.

50 Do you use a specific methodology to assess the training and validation datasets?

Mentions are: ML methodologies and good practices, Statistical analysis, Sampling schemes, Data Management, Train, validation and test sets.

17 Do you rely on a particular set of tools to specify the requirements to define the training and validation datasets?

The majority of answers are that engineers don't use specific methods nor tools to define training and validation datasets. 2 methods are cited once each following ATOS check list and following Stanford ML practices. Some technical tools to split data are mentioned ML platforms (ATOS Codex AI Suite FMLe, MLFlow) with requirement managers (ATOS Checklist for reproducibility), some in-house solutions based on scripts and PowerApps.

31 Do you use specific development tools to build the training and validation datasets?

When asked about their development tools to build the training and validation datasets, many people answer that they use classic methods (without referring to any name, or referring to Stanford) with classic tools (python, sklearn library). Two data engineers mentioned the use of AI to help in the tag phase (AI object detection, active learning and automated annotation tools). Some people have custom tools to display missions' information/images for operational experts' feedbacks.

33 Do you rely on specific engineering tools to integrate training and validation datasets?

Most people say that they don't use engineering tools to integrate the training and validation datasets. Two data engineers use python and in-house pipelines (fisheye correction, viewport projection, training files generation), and one uses DVC to version the datasets. A shared file system is mentioned as a solution to integrate datasets. It is highlighted that the datasets integration requires a platform, a pipeline, and process monitoring tools. The difference between engineering tools (this question) and development tools (question 34) was probably not easy to perceive.

34 Do you rely on specific development tools to integrate training and validation datasets?

Most people don't answer this question, probably because of its similarity with question 33.

51 Do you use specific engineering tools to assess the training and validation datasets?

There is generally no dedicated tool to assess the training and validation datasets. Each partner seems to develop ad hoc solution for this at project level or at enterprise level An IV&V Manager highlights that data management should be improved to better respect needs derived from the system integration.

52 Do you use specific development tools to assess the training and validation datasets?

Same as question 51

3.5. Machine Learning

Summary

The domain of Machine Learning is supported by numerous commercial and academic tools and other technologies (e.g., benchmarks, criteria, methods). Our industrial partners use these technologies for their ML developments, and sometimes add in-house products for

the same purpose. In upstream phases, no methodology seems to be used, but once engineers move to practical implementation, they are empowered by the richness of the current supply. Tools such as Tensorflow, Keras, scikit-learn, Pytorch, are often mentioned, as well as popular benchmarks and criteria. Bottom line, there does not seem to be any complaint about the dependencies over these products.

Details

20 Do you use a specific methodology to select learning methods and algorithms?

No specific methodology used to select learning methods and algorithms. They are chosen among state-of-the-art methods following expert advices.

36 Do you use a specific methodology to design machine learning algorithms?

No methodology for machine learning algorithm design

53 Do you use a specific methodology to assess/benchmark/evaluate your machine learning algorithms?

Some partners use academic well-known benchmarks, criteria and methods (e.g., GLUE for NLP; MOTA, MOTP, IDF, HOTA for Multiple Object Tracking) and other technical criteria (time, CPU load). Sensitivity analysis, risk quantification, formal verification of functional / robustness / safety properties, statistical performance, Fold analysis are mentioned. Many do not have any specific methodology for ML algorithms and re-use/adapt methodologies and criteria from non-ML system evaluation.

21 Do you rely on a specific set of tools to select learning methods and algorithms?

A notebook with classic analysis or several sklearn or Keras/Tensorflow (for example) implementations are tested or other open-source libraries. AutoML and FAST-AI (learning rate detector) are mentioned. It seems that the selection of the model is mainly based on the knowledge of expert.

37 Do you use specific tools to design machine learning algorithms?

Machine learning algorithms are designed with expert knowledge and state-of-the-art on-the-shelf designs on tools like Matlab, Tensorflow, Keras, PyTorch, scikit-learn. One person uses autoML tools.

54 Do you use specific engineering tools to assess/benchmark/evaluate your machine learning algorithms?

The assessment of Machine learning algorithms can rely on a comparison to academic/state-of-the-art models, or to custom benchmarks. The tools mentioned are: Matlab, Python, MLFlow, tensorflow callbacks, keras, Atos Codex AI Suite FastML Engine, Jenkins. Assessment tools for AI specialists will have to be adapted to be used during system integration.

55 Do you use specific development tools to assess/benchmark/evaluate your machine learning algorithms?

Same as Q54. Some use specific python libraries such as motmetrics and trackeval. Some use a HW simulator. Some use in-house tools.

89 Do you use particular engineering tools to deploy your machine learning algorithms?

Isolated mention of: Docker/Docker Compose, MLFlow/Kubeflow.

121 What difficulties do you expect from using mainstream tools (e.g., Tensorflow, Keras, etc.) in your process? Hegemony? Volatility? Certification?…

No difficulties identified/expected by partners until now.

3.6. Knowledge-Based approaches

Summary

Knowledge-based approaches are little used by industrial partners. We received only a few answers on the subject, and most often people do not know about the methods and tools in this domain. We see a tiny number of mentions of ontology tools such as Protégé, some in-house technology, and that's about it. The survey was done during Batch 1 of the program, which is devoted to data-based AI and machine learning solutions. It might be useful to do a second round in 2023 to see if things change significantly since Batches 2 and 3 will include KB use cases and solutions.

Detail

18 What methodologies do you rely on to specify and manage knowledge bases?

Not much activity on this domain for Algo Engineers neither data engineers, they do ML. For Safety Engineers, the situation is the same than for Q16. Neither the less Antidot Technology is cited as a possible tool. In the case of System Engineers, only Airbus answered that there is a trend to standardize a methodology for knowledge bases management.

39 Do you follow a specific methodology for designing knowledge-based AI algorithms?

No-one mentions any methodology for designing Knowledge-Based systems.

56 Do you use a specific methodology to assess/benchmark/evaluate your knowledge-based algorithms?

It seems that knowledge-based algorithms are not used or well-known by questioned industrial partners.

90 Do you use specific methodology to deploy your knowledge-based algorithms to the target environment?

Answers to this question, if any, have not been analyzed yet.

91 Do you use specific methodology to validate the porting of your knowledge-based algorithms in the target environment?

Answers to this question, if any, have not been analyzed yet.

19 What tools do you use to specify and manage knowledge bases?

Very few engineers use knowledge bases, so on all profiles interviewed there are only 3 answers. One data management specialist uses ontology mapping, rdf schema and Protégé software to specify and manage knowledge bases. One software engineer mentions in-house HF base (airbus) and some in-house policy to manage it. The last one listed DSSP: CoR for specification, Antidot and relational database for storage.

Unanswered questions

For all the following questions, the answer was "no" or almost "no". No further detail is given.

40 Do you use particular engineering tools to design knowledge-based AI algorithms?

41 Do you use particular development tools to design knowledge-based AI algorithms?

57 Do you use specific engineering tools to assess/benchmark/evaluate your knowledge-based algorithms?

58 Do you use specific development tools to assess/benchmark/evaluate your knowledge-based algorithms?

92 Do you use particular engineering tools to validate the porting of your knowledge-based algorithms in the target environment?

93 Do you use particular engineering tools to deploy your knowledge-based algorithms to the target environment?

94 Do you use particular development tools to deploy your knowledge-based algorithms to the target environment?

95 Do you use particular development tools to validate the porting of your knowledge-based algorithms in the target environment?

3.7. Human-Computer Interaction

Summary

The subject of HCI is a very rich one. The respondents identify three modes of collaboration between AI systems and humans. In the three modes, the demands are for AI systems to be reliable, reactive, to reduce the worker's cognitive load, to be reproducible, transparent, in short to behave as a trustworthy co-worker. The major risk identified is when an AI system reproduces the same errors over and over again, which corresponds to the famous saying "trust take the stairs up and the elevator down". All these issues are to be taken seriously in the program.

Detail

64 Do you consider some interactions between human operators and AI system?

Nearly all answers are elaborating on human AI interactions, showing evidence of the need to properly consider human AI collaboration. It may cover supervision questions (AI activation/deactivation, overriding AI proposal), anomaly or changes in operational context detection for mutual benefit and performance monitoring, growing trust within partners…

65 If human interactions with your AI-based solution is considered, what do you consider to be the most critical issue to be solved in systems interacting with humans?

The survey shows that, on the one hand, there is a need for AI to be able to give confidence to the users, which could be done by giving guarantees of reproducibility, interpretability and usefulness/relevance of its outputs. It is also expected that the AI does not increase the cognitive load of the human interacting with it, through adequate visualization/presentation of results, low false alarm rate, limitation of the amount of interactions needed. Finally, it is expected that the AI is able to react adequately to the instructions given by the human.

66 What are the main factors that can compromise the trust of human operators interacting with an AI system? (e.g., failures, uninterpretable outputs, …)

The main factors that can compromise trust can be classified as follows:

- information deluge, more complexity

- too many failures, irrelevant results, false positive

- lack of adaptivity and weak reliability of the application

67 Does your problem involve a collaboration between human operators and AI system?

Some industrial use cases do not require any Human AI collaboration, but most of them do. Three modes of interaction are envisaged by the respondents to the survey who answered 'yes'. The first of these modes is the case where the AI produces recommendations to an operator, facilitating her/his task. In a second mode, the AI seeks help from the human when it is no longer able to perform the task alone, and the human can take over or help the system. In a third mode, it is the AI that monitors the human in the execution of a task.

No clear distinction yet between interaction and collaboration.

23 What are the main features requested from a mission/life-critical AI system in terms of interaction with human operators? (Completeness, interpretability…)

Few use case don't imply any human interaction; most do. It is first considered that human shall not be disturb but kept in the control loop while reducing cognitive load, particularly to fix wrong decision by AI assistance. It requires AI to meet requirements such as validity, explainability, accountability, interpretable, reproducibility, completeness, etc. The type of answers that have been collected tends to show a global awareness but still a low maturity about how to globally engineer such solutions.

38 How do you design the (human-machine) interface of machine learning algorithms?

Human machine interfaces for AI functions are not considered differently, UX design applies, at the time without any specific considerations. Which may be inconsistent with previous and following statements about mission/life-critical AI systems consideration, that require specific care.

118 What do you consider to be the best recent achievement in systems interacting with human operators?

Self-driving cars, enhanced surgeon, augmented reality, face recognition, chess and go players are striking examples of AI interacting with human. Within the industrial partners, the main topic is Human AI dialog, with some preliminary results.

3.8. Safety

Summary

The main safety concerns raised by industrials are the potential errors of an AI system and their consequences. Guaranteeing precision and generalization, linearity of inference, characterizing the uncertainty of answers are obvious needs.

Several standards have been designated, depending on the industrial domain and reveal that industrials almost systematically work in standard controlled environment.

Details

123 Do you have applicable safety / system / software / hardware standards (ARP, ISO…)?

In aeronautics (civilian and defense) for FPGA, DO-254, for software, DO-178C, ARP4754 for system level. IEC 61508 or ISO 26262 for terrestrial application (cover FPGA/SW/System) and ISO 21448.

Eurocontrol ED/ARP, DO, ISO, CENELEC, ECSS, STANAG, and other depending on the application domain (IS/IT, Health, Communications, Defense, etc.).

125 According to you, what are the main safety concerns raised by AI?

The main safety concerns raised by AI systems identified by partners are linked to potential errors of AI systems: catastrophic degradation of accuracy, non-linearity of response, uncertainty of output, generalization guarantee, etc.

3.9. Software Interoperability

Summary

The Federative Environment has to be compatible with common infrastructure stacks such as Kubernetes and HPC. Means it must be deployable and runnable on both system (on a virtual machine or an HPC environment) therefore the AI toolchain (such as Keras, TensorFlow, PyTorch, ... for machine-learning dedicated tools) must be compatible with the execution environment regardless of the target infrastructure.

Confidentiality and cybersecurity are two main constraints that should not be neglected when solving interoperability needs. For instance, non-sovereign code should not be used.

Details

128 Do you know any specific constraints on your existing toolchains that could make it difficult to interoperate with the tools of the federative environment in terms of interface?

The federative environment needs to be compatible with mainstream deep-learning framework and to not use US-based code.

One engineer mentioned that they can't use kubernetes on their HPC infrastructure.

129 Do you know any specific constraints on your existing toolchains could make it difficult to interoperate with the tools of the federative environment in terms of tool capabilities?

Cybersecurity and confidentiality due to existing toolchain may be an issue that needs to be addressed.

3.10. Lifecycle Management

Summary

Lifecycle management is an activity that seems to be little addressed in view of the few responses obtained. Sometimes in the course of investigation, the lifecycle approach is generally relegated to a single tool (MLFlow and Git are often used for the model) and sometimes ignored

Details

131 Are you using tools to trace the lifecycle of models and data?

MLFlow, AirFlow, and Git come up regularly, mentioned without precision.

132 Are you applying a methodology to trace the lifecycle of models and data?

Generally left open, a few responses indicate, however, that this approach is still under investigation.

3.11. Requirements engineering

Summary

The requirements are expressed in terms of performance, robustness, explainability, interpretability. When dealing with embedded systems, the usual constraints are mentioned: size, power consumption, real time performance. The methods and tools used for requirements engineering of AI components are the ones currently used by industrials for traditional components.

Detail

1 What are the specific requirements attached to AI components in your dev environment? How do you manage them?

The requirements of an AI component are derived from more global system requirements. They impact different kinds of performances, both software (e.g., inference time, confidence level) and hardware (e.g., accelerator footprint). The main constraint is the size of AI components to run on embedded applications or in real time, or that need to be run on specific infrastructures. The means to control the size of AI components are either metrics or techniques: the metrics mentioned are maximum inference time, maximum number of multiplications/accumulations concerning embedded deployment and the techniques are optimized CNN, Light Architecture (model size), quantization, modification/limitation of frame rate etc. Many non-specific requirements have been mentioned such as using DGA guidelines as a reference. A workbench leader also highlights the importance of a clean code, in particular in case of code generation and a system engineer mentions cyber compliance & security, of the ability to address mobility & asynchronous connectivity and of the notions of trust, efficiency and sustainability.

4 What are the main requirements for mission-critical or life-critical systems interacting with human operators?

Few use cases don't imply any human interaction; most do. Then there is a need for defining rules and/or more precise requirements. Generic rules reflect the intent to keep human in control (leave the final decision to operator) but it is not so well understood yet that this is possible only when the operator is fully aware of the situation at decision time. More detailed requirements still lack accuracy (explainability, robustness, trust, easiness of use, reliability, efficiency, effectiveness) and as such, difficult to turn in clear design constraints.

7 Do you have applicable real-time constraints for your AI system?

Industrial need to take into account time constraints in the order of a few ms to 20 ms, consumption of the order of 1W to 10W depending on the application. Different levels of requirements depending on the application domain

12 Do you have requirements about explicability, interpretability from a methodological point of view?

There are currently some requirements about explicability and interpretability, which is seen important for the future. From a methodological standpoint, is it expected in the following phases: learning, test, exploitation. The expected benefits are: demonstration of the

intended function and innocuity, root cause investigation, debugging, human teaming.
One partner uses explainability for legal purpose to demonstrate the safety of a model.

13 Do you have requirements about explicability, interpretability from a tooling point of view?

The tools for interpretability/explainability could be embedded in a software to be used in production. If explainability is supplied through an external tool, it should provide a user-friendly interface, and the robustness of the tool shall help to efficient V&V of the AI system. It can be noted that explainability/interpretability tools should not be only for model but also for database and data (included in annotation tools).

8 Do you use on a particular methodology to specify your ai-based systems?

The responses to this question show the diversity of approaches. Some have implemented methodologies from or derived from their field of expertise (e.g., aeronautics), others rely on generic methodologies (Agile, Kanban, Scrum, etc.), or on particular frameworks (e.g., VertexAI). In the end, it is agreed that no methodology dedicated to AI is currently used within the consortium members, the methodologies traditionally used by each partner for non-AI systems are used, as far as possible.

10 Do you use a particular methodology to manage your requirements?

No specific methodology used for requirement management of AI based system. The methodologies traditionally used by each industrial partner for non-ai systems are used: RMF, automotive SPICE, INCOSE-->RBE.

9 Do you rely on a particular set of tools to specify your AI-based systems?

The answers from algo engineers clearly stand out from the others. Algo engineers only manage the requirements of the model or the AI application, not the whole systems. So their answers are focused on the requirements. For example, to monitor the size of neural networks they use native pytorch functions (number of MACS, memory footprints, number of weights, etc.) Interestingly, several other engineers answer that they do not have a particular set of tools to specify AI-based systems; common requirements managers are listed such as DOORS, 3DS Requirements Manager, Reqtify, Jira with or without Rmsis, or MBSE tools such as CORE, SCADE.

11 Do you use tools to manage your requirements?

No specific tools used for requirement management of AI based system. The tools traditionally used for non-ai systems are used: DOORS, Capella, Jira, 3DX, Excel. Two tools clearly stand out for every profile that answered: DOORS (IBM requirements managers) and Excel spreadsheets. It is worth mentioning that system engineers use the same tools for requirements management and for specification of AI-based system.

3.12. Design

Summary

This section examines the use of formal methods, processes and methodologies and the supporting tools in the design of AI components and systems. There is little use of such approaches in current practice, and consequently there is an obvious need for Confiance.ai and other parties to deliver guidelines and tools for this purpose. When companies integrate components from third parties (i.e., suppliers) they tend to ask for guarantees of some sort in the contracts.

Detail

24 Are you already using formal methods in your traditional developments?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers

There is no use of formal methods (only exploration in research project)

29 Do you use a particular process to design your AI components taking into account system requirements?

This response received very few replies and it is difficult to find in these replies the existence of a process to meet the specifications. Only the use of AutoML seems worth mentioning. A lack is identified by industrial partners on that point. Strong expectation from new standards (e.g., Eurocae)

25 Do you integrate AI components coming from third-party providers? If YES, how do you estimate the trustworthiness those components?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers

AI components from third-party providers are often integrated in AI based system. The trustworthiness of those components is estimated by either evaluation of AI components on customer internal dataset or qualification of AI components on reference dataset.

Furthermore, some industrial partners consider having a more contractual approach with third-party providers, including 1/ explainability tool that can be run outside the ML (no access to source code) 2/ explainability/data/process requirements to be passed to the third party 3/ to understand the guarantees that we can take credit from (black box)

26 Do you employ a specific methodology to design your AI-based systems or services?

The responses mention the use of generic tools (Agile, Kanban), of the Stanford ML methodology and of the A3/WFR.

27 Do you rely on particular engineering tools to design your AI-based systems or services?

Answers to this question, if any, have not been analyzed yet.

28 Do you rely on particular development tools to design your AI-based systems or services?

Answers to this question, if any, have not been analyzed yet.

3.13. Algorithmic Engineering

Summary

Too few details are provided to extract any information other than the non-use of algorithmic engineering in the AI process.

Details

42 Are you using Algorithmic Engineering? How do you manage it?

Mostly not. Sometimes there is some kind of engineering to fit the requirements but little detail is given. I suspect that this question did not receive proper answers.

43 Are you using Algorithmic Engineering tools?

None.

44 Are you using Algorithmic Engineering methodology?

There is no mention of algorithmic engineering methodology.

3.14. Assessment

Summary

In general, the assessment method of AI-based systems is partly based on the application case: while there are indeed criteria for evaluating individual AI components of a system, the integration of humans or the hybridization of the two proves less covered.

There is definitely a lack of methodology for characterizing an AI system, which currently seems to be done on a case-by-case basis, sometimes in line with a certification process, but very often in an empirical way with a clear confusion between system performance and trustworthiness.

As it stands, the decision not to integrate AI into a solution is based on the fact that the IVVQ or criticality criteria of the function (with respect to the DAL) underlies a risk that the added value cannot compensate.

Details

45 What are the criteria used to evaluate the performance of a mission/life-critical system relying on AI systems and human operators' interactions?

Some are not concerned with mission/life-critical systems. For the others, this question is not always tackled at AI component level (only), but several levels like module, component, full system, end user testing. This is highly dependent on the application case. When it comes to performance evaluation, we have numerous examples of criteria for ML components, a few examples of criteria for operators' interactions, but the combination of both is still weak or under evaluation.

46 Do you have any criteria to characterize an AI component?

This question has been understood from different standpoints. Some criteria are defined in Autonomous Rail Transportation domain (GoA) or in Tacking Accuracy Measurement domain (MOTA, MOTP, IDF, HOTA, CMC, ...). For other domain, criteria seem to be defined ad-hoc for each project:

- validations steps of the ML component (training, sub-component, whole component)

- level of autonomy, or level of automation

- type of AI

- criticality in the application function

- ML metrics, criteria for Tracking Accuracy Measurement, etc.

47 Would you consider an empirical (probabilistic/statistical) thorough assessment of AI components to demonstrate their trustworthiness?

There is a need for clarification between assessing performance on one side and showing evidence of integrity on the other… Probabilistic/statistical thorough assessment of AI component is considered by Industrial partners, but not in all cases. Furthermore, this raises some questions:

- representativeness of validation dataset?

- how to justify/quantify the credit taken by ML component massive testing evaluation?

- how to justify the relatively small amount proof test/verification at system level?

48 What are the major criteria leading to adopt / reject AI solutions, from the methodological point of view?

If related to 'specification' or 'design' phases, the question would have referred to the cost vs performance ratio of the insertion of an AI module in a solution. Here, for 'Evaluation' purpose, the answers focus on data availability and quality, in relation with the level of criticality of the AI-based function.

49 What are the major criteria leading to adopt / reject AI solutions, from the solution point of view?

- IVVQ criteria (performance…),

- or criticality of the function (=> DAL level), that could not be covered by AI solution at the time.

- trade-off between additional value provided by AI solution vs industrial risk of changing the "traditional" solution.

3.15. Deployment

Summary

This section mainly addresses deployment for embedded systems, since deployment on large-scale infrastructures or cloud do not seem to be a subject of too much concern. Most applications seem to be standalone on embedded hardware without connection to a "home-base" cloud. Therefore, edge computing is not mentioned as a need for the moment, although some might think that it will be the case in the near future. In terms of hardware targets, there is a great diversity from CPUs to GPUs including manycore, ASICs etc. Engineers use the tools and methods supplied by hardware manufacturers to port their AI components to the embedded platforms, but there are several associated weaknesses in terms of optimization, resource usage, energy consumption and performance estimation, also because the AI development platforms used do not provide such services. There is a lot of room for improvement on the deployment phase, which is one of the focuses of EC7. This

project highlights the choice of the N2D2 tools which supports implementations on several target platforms.

Detail

70 What is the typical usage mode and calculation workload of your AI-enabled solution? 24/24? On demand? level of solicitation (high/ moderate/ low)? Permanent/periodic/ at times? If YES what are the means (method, tools) used to prevent and face adversarial attacks on your AI components (bricks)?

Typical usage mode is permanent (AI system always running when global system is operational).

In terms of algorithm activation, continuous or on-demand invocation of AI algorithms. Confirmation of the need for algorithm scheduling, segregation to allow several algorithms to coexist (spatial and temporal isolation of algorithms) and system monitoring.

71 Could you provide details about the typical resources available for the AI system execution platform (memory, core, power consumption…)?

Need to take into account time constraints (of the order of a few ms to 20 ms) Consumption of the order of 1W to 10W depending on the application Different levels of requirements depending on the application domain

72 Could you provide details about typical target environment for the AI system? Cloud computing platform? Embedded platform? Other?

No consensus on the choice of platform in the industrial responses: CPU, GPU, manycore, ASIC, AI accelerators. EC7 Comment: This is consistent with the approach of the N2D2 tool which provides a tool for mutualizing the implementation of AI algos on different platforms.

The subject of the coarse-grained estimation of resources before the implementation of algorithms seems relevant. It is also a question of defining these KPIs.

73 Could you precise on what types of embedded targets would you possibly deploy AI components? Mono-core, multi-core, many-core CPus, FPGA, DSP, other accelerators? Conversely, is there any target that you exclude a priori (e.g., GPUs, FPGAs).

Same answer as Q72

74 Is your system embedded (e.g., in a car)? If yes, would you consider deploying part of (or all) of the AI functions on cloud platforms? How could the embedded system and the cloud computation platform be connected? (5G, IoT network…)

Mainly embedded, with sometime AI deployed on cloud platform.

In terms of activation of algorithms, continuous or on-demand invocation of AI.

Edge computing algorithms: There is little need for edge computing. Generally, the embedded target is not connected to the cloud, all the calculations are done at the level of the embedded platform.

75 Do you use specific methodology to port your AI-based systems or services to the target environment?

Little or no methodology for embedding AI algorithms. Today the operation is in a back-and-forth mode between the algorithm designer and the teams in charge of implementing. There are few KPIs to systematically instrument a methodology. No global end-to-end approach taking into account the constraints of embarkability. No projections on the estimation of resources and on functional performance according to available resources

EC7 comment: the tools today (especially open source) do not allow a fine configuration of the implementation of algorithms. There are many abstractions that are made, making the implementation of the algorithm rather opaque.

76 Do you use specific methodology to validate the porting of your AI-based systems or services in the target environment?

Same answer as Q75

77 Do you use specific engineering tools to port you AI-based systems or services to the target environment?

Algorithm engineers use methods such as engineering deployment pipelines and usual packaging (conversion to embedded format for example) to port AI based systems to the target environment. Those tools are provided either by cloud platforms, HW manufacturer ownership modules, or in-house tools. The same tools are mentioned by SW and system engineers: Docker, TensorRT.

78 Do you use specific engineering tools to deploy you AI-based systems or services?

Engineers use the same tools (engineering or development phase): cloud platform tools, HW manufacturer, Docker compose. MLFlow and CUDA for GPUs. The difference between the AI-based system (this question) and the algorithm (question 89) was probably not easy to perceive, and the same fits for question about tools in development phase vs tools in engineering phase.

79 Do you use specific development tools to port you AI-based systems or services to the target environment?

Not a lot of profiles port AI-based system to target environments (not many answers). The tools mentioned are cloud platform tools, PIP, Container (Docker).

80 Do you use specific development tools to validate the porting of your AI-based systems or services in the target environment?

No dedicated tools. Standard ones mentioned (Cloud tools, Jupyter Notebook)

81 Do you use specific solutions to port you AI-based systems or services to the target environment?

As for the platforms, no consensus on the choice of a framework: mention of TensorRT, CUDA and other standard tools. Little feedback from industry on this subject.

82 Do you use specific solutions to validate the porting of your AI-based systems or services in the target environment?

83 Do you use specific methodology to port your machine learning algorithms to the target environment?

84 Do you use specific methodology to validate the porting of your machine learning algorithms in the target environment?

For these three questions:

Embeddability needs are specific to each application domain (a lot of heterogeneity in real time needs, resources…). Nevertheless, the topics of resource estimation and performance projection are underlying needs in the answers of the industrialists.

There is no mention of the topics of algorithms optimization (pruning, quantization…).), which is quite surprising considering the complexity of embedding DNN algorithms

No reference hardware platform to embed AI algorithms

No reference framework, a clear need to better master some implementation steps (need to concretize abstractions made by popular frameworks)

The ability to estimate the necessary resources should help in the selection process of hardware platforms

85 Do you use particular engineering tools to port your machine learning algorithms to the target environment?

Same answers as question 79

86 Do you use particular engineering tools to validate the porting of your machine learning algorithms in the target environment?

No dedicated tools have been mentioned.

87 Do you use particular development tools to port your machine learning algorithms to the target environment?

No dedicated tools have been mentioned.

88 EC7 #Deployment Do you use particular development tools to validate the porting of your machine learning algorithms in the target environment?

No dedicated tools have been mentioned.

3.16. Supervision/Monitoring

Summary

Generally speaking, the process of integrating the monitoring of AI-based systems does not seem mature. Various open-source tools are frequently mentioned (MLFlow, Kubernetes, Tensorboard, Prometheus, Grafana) as well as tools delivered with AI platforms (Azure, AWS Sagemaker), sometimes in-house tools monitoring specific system indicators are mentioned.

It is quite clear that no formal monitoring methodology is currently applied.

Details

103 Do you implement supervision of your AI-based systems?

AI component supervision or AI system monitoring are not implemented yet, at least not in a specific way, particularly suited to ML. We have to think about it!

106 How do you supervised your AI based application?

108 Do you use a specific methodology to monitor/track/supervise your AI based systems or services?

109 Do you rely on specific engineering tools to monitor/track/supervise your AI based systems or services?

110 Do you rely on specific development tools to monitor/track/supervise your AI based systems or services?

Engineers use a variety of tools to monitor their AI based systems, some mentioned open sources tools (MLFlow, Kubernetes, Tensorboard) or tools provided by Azure Cloud Platform. Spotifre has been mentioned for potential future use.

Some mentioned in-house tools, workflow and interface (KPIs, tagging interface, end-user vetting). It is worth mentioning the versioning of model is mostly done with MLFlow.

The use of additional module to double check the result of the task has been mentioned too.

110 Do you rely on specific solutions to monitor/track/supervise your AI-based systems or services?

Four tools have been mentioned to monitor and/or supervise AI-based systems: MLFlow, AWS Sagemaker, Prometheus for cloud and Grafrana for hybrid approaches.

3.17. Certification

Summary

It seems clear to everyone that certification of AI component will have a major impact on the current practices, for methodologies as well as for tools used in the development chain, essentially because there is little or no use of certification tools and methods for AI systems at the moment. Engineers are keen to apply such changes if the added value is demonstrated. However, they ask to have some freedom of choice, not to be constrained by the requirements of certification tools and methods. Assurance cases are not known or considered not applicable.

Detail

97 Modifications of existing engineering (safety, design, implementation…) and certification practices seem to be necessary to integrate AI components. According to you, what will be the practices most impacted by AI from a methodological point of view? And what are the practices that you would like to be the less impacted (if you had the choice)?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers.

Identified as most AI-impacted engineering and certification practices: Data engineering and management through the life cycle (incl. Data analytics).

Some that are expected to be the less impacted: all systems layers, including safety; SW w/o ML item but integrated with ML item

Some new practices expected: Data design and V&V; ML item design, implementation and V&V; Toolchain Data-ML item-System simulation (ML model validation); Toolchain Data-ML item - HW (ML implementation)

98 Modifications of existing engineering (safety, design, implementation…) and certification practices seem to be necessary to integrate AI components. According to you, what will be the tools most impacted by AI? And what are the tools that you would like to be the less impacted (if you had the choice)?

Warning: this analysis, done by EC6, only concerns IVV engineers' answers

Identified as most AI-impacted tools: data configuration management tools, modeling & simulation tools, test tools, CI/CD tools.

Some new tools expected: data design and V&V tools; new tools for ML item design, implementation and V&V; toolchain Data-ML item-System simulation (ML model validation); toolchain Data-ML item - HW (ML implementation)

101 Do you use a specific methodology to certify your AI-based systems or services?

There is no deployed methodology yet to certify AI-based solutions or services. As reduced as possible impact is expected. Eurocae/SAE have started showing the way, Confiance.ai should provide the method and tools. However, some partners prefer that no specific methodology becomes required for AI certification so that they have freedom in the choice of their methodology.

102 Do you use specific engineering tools to certify your AI-based systems or services?

Algorithm engineers are ready to adapt their tools and algorithms to be compliant with certification requirements. Then Confiance.ai tools and methodologies are of great importance to satisfy this need.SEPConcerning System IVV engineers, they already think about specific tools, for instance Assurance Case. But they ask that this tool should not become mandatory but remain as a possible choice for an applicant. This need of freedom is interesting to note.

105 Do you use development tools to certify your AI-based systems or services?

No specific engineering or development tools to certify AI-based systems or services. Each partner seems to develop ad hoc solution for this at AI component level (Python / Jupyter libs), without changes in System V&V toolchain used for certification.

100 Would you consider redeveloping/refactoring some components of the AI workflow (framework, libraries, tools) in order to make them compliant with certification objectives?

Algorithm engineers are ready to adapt their process and toolkit to be compliant with certification requirements. More globally, there is a cost and time issue here to be solved.

107 Do you know the approach of "assurance cases"?

No knowledge about assurance case or (for one partner) assurance cases are considered not applicable.

99 Would you accept drastic changes in your certification practices in order to introduce AI?

This question received only a small number of answers. It seems that many are not ready to accept drastic changes. Those who seem to accept drastic changes will apply criteria in order to move, such as certification needs, costs, scope.

3.18. Other questions

Summary

This section gathers questions that address complementary aspects of the program and do not fit in the previous sections.

- which engineering phase raises more challenges for AI? The answer is "all'.

- what criteria for adopting AI solutions: more or less all criteria one can think of.

- how to assess statistical performance: same as for conventional components

- quality assessment for annotations: double labelling with sometimes a tool to help

- adversarial attacks and forensics: no solution used as of now.

Detail

119 According to you, what is the engineering phase (lifecycle) that raises the more challenges and/or missing answers from industrial point of view?

All engineering phases in AI development are raising concerns, particularly when it comes to safety considerations. But Specification, Verification & Validation, and Certification are the most adduced.

Note also three relevant concerns (crossing phases):

- manage embedded AI

- manage functional and non-functional performance over time

- secure human / AI / System relationship and its management within the engineering team

96 Do you use criteria to adopt / reject AI solutions?

For "non-Thales" interviews: KPI such as performance, effectiveness, confidence, robustness, cost for the user, computational aspects (memory usage), as well as user acceptance, interpretability

For Thales interviews: qualification/certification; cost reduction; trustworthiness (confidence and robustness); interpretability, usability, and computational aspects); however, several Thales respondents answered "no criteria".

117 If you are already using AI, what are the main technologies (programming languages, frameworks…) used in your workflow?

Python, openCV, Tensorflow, Keras, Pytorch, C++, GTF/EVE psybernetixs framework, Unity3D/MLAgents, GYM (openIA), gaml/libsvm framework are the main technologies used today.Note a mention of C++ for embedded AI, and a reference to COMBI, a Human Machine Teaming Thales framework to translate human operational intentions in technical parameters of the machine, and vice versa.

124 The performance of some AI techniques, such as Machine Learning, can only be characterized in a statistical manner. Have you already faced the same situation for non-ML systems (e.g., GPS, Kalman…). If yes, how do you "guarantee" their performances? Do you think that such analogy could help addressing some of the problems raised by AI?

It is not something new for algorithm engineers to have to characterize an algorithm in statistical manner. They cite different technics as massive Monte Carlo testing for example. Then statistical performance assessment is not considered as a blocking point. Algo Engineers insist more on the need of explainability and operational domain representation.For Safety Engineers asking such a statistical demonstration is not that frequent, but yes it can happen. For example, FANS B Performance demonstration show compliance to ED120, ED106, ED89, ED85. For instance, "a maximum of 1 message corrupted by the Airborne system over 10 000 messages exchanged". Similar approach will establish the performance of AI, but not its integrity.

130 Are you using Model Based Engineering practices? If YES, could you provide details about those interoperability constraints in terms of tool capabilities

Warning: this analysis, done by EC6, only concerns IVV engineers' answers

Model Based Engineering is already used in Aeronautical domain (but not in other domains. Some tools are used, such as Arcadia/Capella, and MBSA is expected to be used in the future.

140 Do you use a quality assessment process to qualify your annotations?

Very few answers from Algo Engineers. It seems that annotation quality is a concern when another company performs the annotations work. Some Data Engineers use peer review or try to determine automatically possible error observing manually the failure cases of a trained model.

For IVV engineers, no specific quality assessment process detailed for dataset annotation qualification.It seems, Safety Engineers did not properly understand the question. To do data annotation checking the CVAT is used by ATOS in combination with visual checking. Partners perform double labelling of data

69 Can your AI-based system be subject to adversarial attack? If YES, what are the information required to perform this analysis?

This is a concern for most, but considered out of reach at the moment. It seems that adversarial attack defenses methods are only available in academic labs yet.

68 Do you need means to perform post-mortem (forensic) analysis? If YES, could you precise how the responsibilities are shared among those two components?

Forensic analysis is not required in all use cases. If relevant (Post Mortem (forensic) analysis is performed in aeronautical and automobile domains), it may be for a better understanding of the system design (identification of corner/extreme cases, enhanced training data sets, etc.), but mostly to comply with safety regulation (in case of failure, context, traceability & explainability of a decision mainly). It is important there to consider compliance with GDPR.

4. Annexes

4.1. Specific analysis done by EC6 on IVVQ aspects by profile

Introduction

This document presents the analysis of industrial questionnaires with regard to EC6 concerns. For the analysis, we have analyzed the different questionnaires going from data engineering to system IVV engineering. We mainly focus on reviewing the questions related to EC6 concerns and not in relation with EC1 or EC2. In the following section, we provide per engineering branch, a summary indicating the topics covered, the general trends expressed by the industrialists and the singular points and impact on the project.

Synthesis

Data Engineering

The artificial intelligence based systems (AIBS) engineering go throughout different phases, the data engineers focus mainly on the validation & verification, development and the specification phases despite the ones of deployment and certification qualification. The data engineers are suffering from the lack of tools and methodologists at some phases; they just use representative data coming from raw data to specify the requirements of their systems a substitute of the ODD. The AIBS design lacks the means to formally specify the requirements of the systems in terms of completeness, reliability and robustness. Another feature to consider while designing AIBS is to consider the human in the loop while the confidence of the AI component is too low to validate an action. The paradigm of model-based system engineering is not used, this fact opens new research questions in this field. However, we have some initiatives to trace the lifecycle of models and data, examples of tools that can be used: Atos Codex, MLFlow, Picsell…

DM-PT (Data Management)

There is no particular methodology to characterize/design the operational domain.

No means to perform post-mortem (forensic) analysis are needed because they might not be compliant with GDPR rules.

No specific methodology followed for the approach of "assurance cases". However, it is planned that in the first phase, the AI will only be a prediction that will need confirmation by human in order to monitor the rate of good prediction, increase the datasets of human-labelled data and improve the model performance over time. For removal of sensitive data, we also can use a proxy task: number of sensitive information find daily/weekly in order to detect a potential drift in rate of findings which could indicate that data might have changed, calling for a potential revision of the classification algorithm.

Some industrial partners monitor, track, supervise their AI based systems or services by the means of periodic tests (daily and weekly tests).

Algorithm Engineering

There is no consensus on the methodology to design the ODD, but it is not addressed at the algorithm engineer level. However, in specification steps, prior knowledge of the operational domain is integrated from taxonomies and dictionnaries.

In terms of HMI for life-critical AI the focus is essentially on assistance in decision to reduce cognitive load. Additionally, a human-based validation relies heavily on interpretability and robustness properties.

The specification process should be impacted by AI, with the description of functions at higher levels, along with more detailed response boundaries, as well as a more complete expression of the needs in terms of V&V of the datasets, thresholds on statistical metrics, and applicability of formal verification methods.

The validation of AI deployment yet rely mainly on internal standard engineering pipelines and cloud platforms, with internal methodologies and tools and with traceability of models and data through git or ML flow. The V&V is expected to be more complex for AI-based systems, and should include additional proofs and tools, targeting specific properties like explainability, stability (robustness), or representativity. Moreover, internal development should be prefered over third-party tools and libraries, to avoid the complex problem of tool certification.

The supervision and monitoring of AI-based systems currently rely primarily on qualitative and quantitative metrics and errors, as well as operational monitoring to prevent data corruption issues (OOD detection).

AI-based implies the need of new types of tools for the whole ML lifecycle, for database management, for data preprocessing and analysis, for meta-learning and speedup of hyperparameters search, for deployment configuration ensuring reproductibility, for model verification, both statistical and formal, as well as specific tools for RL, for scenario management, online data analysis and reward design. On the other hand, no change is expected for low-level implementation tools, typically for FPGA.

Finally, the focus of the Confiance.AI project should be on V&V and quality, with robustness as the main safety concern, as well as confidence.

System Engineering

There is no specific methodology used to characterize or design operational domain, however mock-up HMI tools may be used to define operational scenarios.

The main safety concerns raised by AI are related to uncertainty in output, domain drift, and their poor generalization capabilities.

As analysis methods, envelop justification may be used at design stage, formal methods are used sometimes at software verification stage, and post-morterm (forensic analysis) are used at operation stage to provide explainability at system level by translating the ML item output in an understandable way. The latter analysis is not compatible with GDPR rules to be generalized to other system engineering needs.

Safety Engineering

Only 3 partners have answered the survey related to Safety Engineer point of view: Thalès (Railway), Thales (Other), and Airbus.

None of them already AI for industrial applications. They only answer with exploration or Proof of Concepts point of views. In consequence, no industrial methodology and tool dedicated to AI do not exist yet in their current process. They still refer to sectorial standards which are not adapted to AI-technology. It is the same for engineering tools: they refer to current tools: DOORS, Capella, Matlab-Simulink, etc.

Furthermore, it is quite surprising that they don't seem familiar with the notion of "Assurance Case", whereas Safety Case is a current practice in the industry.

In conclusion, there are few information to extract from this interview. We can only consider the interest of the partners for Confiance.AI program. From the point of view of a Safety Engineer, the first result is about methodologies more than tools (we understand that there is no existing global toolchain in their current process).

Knowledge Engineering

No responses

Software Engineering

Particular methodologies to characterize or design the operational domain can include Tasks Modelling, but are generally inexistent or not addressed at the AI engineering level.

A mission/life-critical AI system (in interaction with human operators) should be able to provide:

Humanly interpretable confidence scores on its results,

Humanly interpretable errors (due to wrong inputs, for example),

Humanly interpretable additional information (to reassure human operators).

The use of formal methods in traditional developments seem to be limited to project management strategies for such as Agile Methods and the use of DevOps infrastructures for continuous integrations which intend to mitigate bug probabilities, allowing for example automatic tests and fast fixes in urgent situations.

AI techniques could to come in support of development or verification activities (e.g., automatic testing) more specifically in the case of monitoring; however, in other cases it is neither clear nor expected.

Concerning the integration of AI components from third-party providers, it seems to be rare or inexistent (probably because the trustworthiness of it would be hard to estimate).

Post-mortem (forensic) analysis are mandatory for legal reasons for specific events (being able to "replay" the event according to regulations, to discriminate what comes from Human error or from the systems) in specific industries such as aeronautics. Some other industries seem to discard this possibility for apparently not needing it or for incompatibility with GDPR rules.

There are no identified specific solutions nor methodologies to validate the porting of knowledge-based algorithms in target environments.

In order to trace the lifecycle of models and data, there are no identified methodologies, but tools such as MLFlow (for models) and DLC.

The supervision and monitoring of AI-based applications is possible through platforms such as the AWS SageMaker Monitor. Intrinsically, the aeronautics industry has all the methodologies to monitor/track/supervise in operations. The question is how these methodologies need to evolve because of the AI item. Supervision by the machine and by the user are both considered. Thus, the human failure needs to be further investigated, detailing how the user contributes to the contextualization of a misbehavior of the model (i.e. collaborative systems, linked to explainability and user center design).

Concerning all parts of the process, the qualification of the tools (embedding models learnt from these tools) is an effort very expected from the confidence point of view.

Finally, assessment lists for AI such as ALTAI (Assessment List for Trustworthiness AI) are not used (yet).

Data Engineering

In early stages of AI integration, validation should be perfomed with the human in the loop, for prediction success monitoring, for additional labellisation and model performance improvement. In terms of supervision, monitoring such as distribution shift of sensitive data ratio should allow revision of the classification algorithm. For now, this monitoring is mainly done through periodic testing.

Human Cog. Engineering

Means, such as data logs of inputs, outputs, user and context to perform post-mortem (forensic) analysis are needed.

There is no specific methodology to monitor, track, supervise th AI based systems or services of the industrial partners.

For some partners, the main safety concern raised by AI is the complacency (over-reliance when system is wrong). As with human teammates, we need to be able to detect when something is wrong and be able to accomodate to it.

In order to trace the lifecycle of models and data, gitlab is used.

System IVV Engineering

The answer provided by the System IVV engineers are numerous: more than 75 questions answered by one or several partners. All texts have been used in section 3 of the main report.

WB Engineering

In order to identify and characterize operational domain, human factors are used and comparisons are done with real data.

AI technics are considered for verification activities

- Unsupervised Learning for tests monitoring,
- Reinforcement Learning for identifying system failure cases.

Post-mortem analysis may be needed in order to improve AI algorithms and identify corner/extreme cases.

Qualification activities depends on the limited amount of data available in order to define the necessary training and validation datasets.

The main safety concerns raised by AI will be the ability of AI to provide a level of safety compared with current standards.

In order to trace the lifecycle of models and data, Windchill (configuration management tool) may be used.

Conclusion

We collect very few answers per engineering but the System IVV engineering. We can draw some high-level conclusions for the analysis of the industrial partners about EC6 concerns.

The integration of AI components on industrial systems are very limited and basically focused on perception-based system using a data-based AI. Hence, no industrial systematic methodology and tool dedicated to AI do not exist yet in their current process.

Mainly performance, explainabilty and robustness criteria are currently attached to AI components choices. So, industrials do not expect a great impact in their current practices regarding IVVQ in general. The methodologies traditionally used by each industrial partner for non-AI systems are expected to still be used for AI systems except for the data and ML engineerings for which high expectations have been raised – at methodological, tool-

support, standard levels. Those traditional methodologies rely on conformance to prescriptive recommendations from standards while alternative qualification/certification approaches like assurance case-based are not in the customs.

4.2. The questionnaire