Franco-German position paper on "Speeding up industrial AI and trustworthiness"

Julien Chiaroni, Sonja Zillner, Natalie Bertels, Patrick Bezombes, Yannick Bonhomme, Habiboulaye Amadou-Boubacar, Loic Cantat, Gabriella Cattaneo, Lionel Cordesse, Edward Curry, et al.

FRANCO-GERMAN POSITION PAPER

ON "SPEEDING UP INDUSTRIAL AI AND TRUSTWORTHINESS"

Consultation Version

May 2021

This position paper summarises the presentations and discussions of Franco-German workshops, including numerous received feedbacks. In addition, it aligns with national and European related roadmaps to industrial and trustworthy AI.
It should be seen as a first step in speeding up Industrial AI and AI trustworthiness across Europe! Open for Europe!

The full benefit from using AI to generate value for businesses, societal wellbeing and the environment is still to be fully realised. To lower adoption barriers of Industrial AI, **challenges on multiple levels** (technical complexity, trustworthiness, industrialisation, data frameworks and infrastructures, etc.) need to be addressed to benefit from its full socio-economic potential for economy, society and welfare.

It is now time to foster the development of "industrial and trustworthy AI" and nurture European innovation and sovereignty ambitions and benefit society and leading European industries.

In this position paper, a **comprehensive *industrial and trustworthy AI framework* that clusters the priority area for AI research, innovation and deployment is introduced.** It covers **tools and methodologies** that support the design, test, validation, verification, and maintainability of AI-based functions and systems and addresses the **development of AI-based process and systems** to demonstrate its integration into new products and services.

Conformity assessment schemes, balancing innovation, business and European perspectives, are considered to connect risk management, functional and trustworthiness requirements to industrial processes. In addition, adequate **standards supporting industrial AI and trustworthiness** will play a central role.

Implementing the *industrial and trustworthy AI framework* will require resources beyond the means of any European private stakeholders. Therefore, strong support from ecosystems, governments, and Europe might not be an option but necessary.

AI WILL PROFOUNDLY IMPACT THE EUROPEAN ECONOMY, SECURITY AND SOCIETY

Based on ongoing surveys, IDC predicts that global spending on Artificial intelligence (AI) will increase from 40 billion euros in 2019 to 119 billion euros in 2025. Thus, AI, as a key enabling technology, will substantially increase competitiveness in all industrial sectors of the European economy (automotive, aeronautics, health, industry 4.0, energy, environment, waste management, agriculture, *etc.*), as well as impact the entire engineering process, from design and manufacturing, leading to new products and services portfolio and participating to the economic recovery of economic sectors. AI will increase the production line's automation capabilities, optimise the maintenance for the manufacturing sectors, and, ultimately, develop its resilience and sometimes scale-up, which importance is highlighted by the current covid-19 crisis. Finally, AI is more than ever a critical technology for both European digital sovereignty and competitiveness in a context of intense international competition.

Furthermore, AI will be an essential part to create new businesses in and for Europe! Business impact, AI, getting value out of data, data spaces, data infrastructure and connectivity besides security and other technologies such as microelectronics will be essential, including the domain knowledge to achieve a data economy in and for Europe!

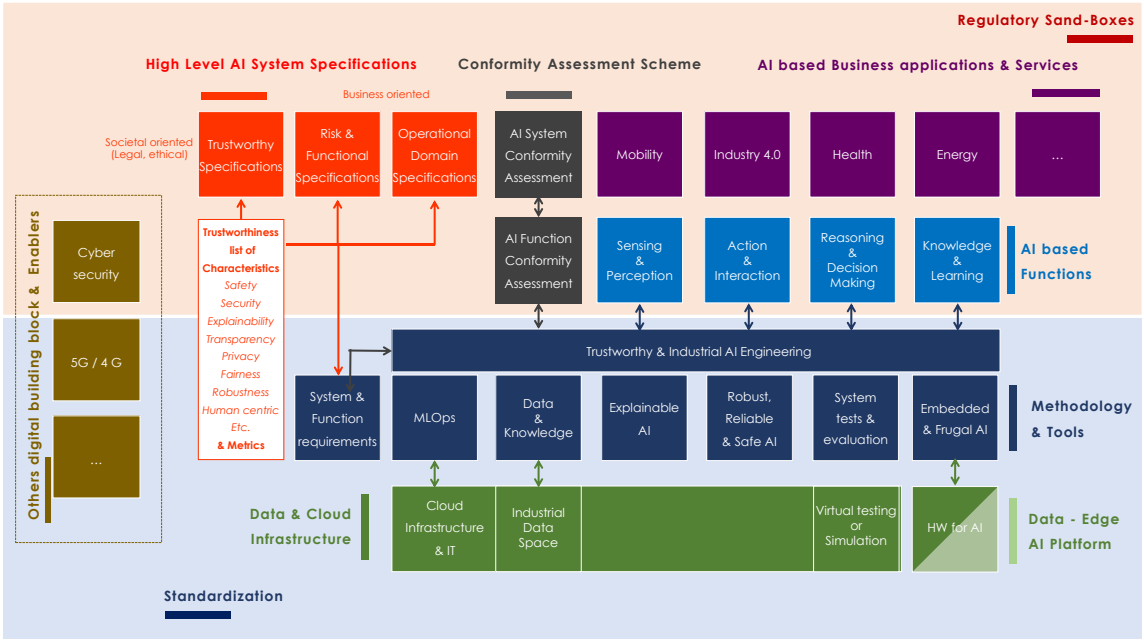
AN *INDUSTRIAL AND TRUSTWORTHY AI FRAMEWORK TO* BOOST INDUSTRIAL COMPETITIVENESS AND PROMOTE EUROPEAN VALUES

AI faces challenges on different levels that need to be addressed to speed up the adoption of industrial AI and trustworthiness. A successful strategy to overcome these challenges requires collective actions from all stakeholders around the objectives of a common Industrial and trustworthy AI strategy to strengthen synergies and develop best practices.

Based on the input received through several workshops, projects, roadmaps and consultations, a comprehensive ***industrial and trustworthy AI framework*** clustering the priority areas for research, innovation and deployment have been consolidated (see Fig x). A wide range of (interdisciplinary) activities needs to be fostered to develop a complete standardisation and technical framework.

This framework that combines national efforts with European existing frameworks and roadmaps is meant to be very instrumental in **fostering the development and speeding up the adoption of Industrial and Trustworthy AI** in Europe, strongly contributing to the current efforts and ambitions

of BDVA[1], Gaia-X[2] and other initiatives, such as microelectronics, of creating a solid Data and AI ecosystem in Europe and taking benefits of a national initiative such as in France with Le Grand Défi[3] and in Germany with Plattform Lernende System[4]. But also in all other countries are significant initiatives to cooperate with!



INDUSTRIAL & TRUSTWORTHY AI ENGINEERING

The framework envisions the development of a **set of tools and methodology for trustworthy and industrial AI** (blue stack) that will support the design, test, validation, verification, and maintainability of AI-based systems in conformity with all the specifications (orange stack: operational design domain, risk, functional and trustworthiness), cooperating with European regulation.

To speed up Europe Industrial AI deployment in leading European domains, we must have an "adequate" set of tools (software, etc.) and methods that support trustworthy AI at the component and system level, wherever the specifications come from regulation, societal concerns, safety or security, *etc.*

---

[1] *An industry-driven international not–for-profit organisation with more than 200 members all over Europe that aim at positioning Europe as the world leader in the creation of Big Data Value and data-driven Artificial intelligence (www.bdva.eu)*

[2] *An international team stemming from several hundred organisations working together to create the next generation of data infrastructure (www.gaia-x.eu)*

[3] *A strategic initiative of the French innovation council that aim at promoting and funding disruptive innovation on trustworthy AI for industry*

[4] *German's Platform for Artificial Intelligence https://www.plattform-lernende-systeme.de/home-en.html*

To achieve this goal, we must revisit "classic" engineering (algorithmic engineering, software engineering and system engineering) to ensure the system's compliance with requirements and constraints. New sets of methods and tools shall be developed to streamline AI systems life-cycle phases, from data collection and knowledge engineering to design and in-operation monitoring and maintenance. The industrial challenge is then to design end-to-end the entire "AI system engineering" process covering the entire value chain, thus making it possible to industrialise AI.

This requires targeting an implementation on both private infrastructure and Gaia-X cloud type infrastructure (Green Stack) and consider the needs of edge and embedded AI solutions. Hence, it will create synergies between industrial data space, computing capabilities, and AI system engineering while developing the competitiveness of strategic European industrial data-driven value chains - where we can build on BDVA, Gaia-X and other initiatives in Europe!

France and Germany have already launched projects on these topics, such as *confiance.ai* (within the "Grand Défi" in France), *deel* programs, the *platform learning system*, *MIAI-Embedded and distributed AI and hardware architecture for AI* or the project *KI Absicherung*, paving the way for a holistic joint Industrial AI approach and scale-up at the European level.

AI-BASED FUNCTIONS AND SYSTEMS

The framework promotes the **development of AI-based functions and systems** (light blue stack) and demonstrates its integration into new products and services (purple stack). Today, European industrial domains are suffering from a lack of ambitious industrial project of AI-based functions and systems that demonstrate value creation with AI. It partially explains the difficult scale-up from proof of concept to new industrial products and services that can federate a strong innovation ecosystem. The set of tools and methodology for industrial and trustworthy AI (blue stack) will support these developments and allow the industry to speed up AI industrialization.

Industrial projects are already ongoing in France and Germany and across Europe, for example, in Mobility, Industrie 4.0, Health or Energy. As cooperation for "close to market product and service development" could be challenging for competing European industries, it is proposed to focus cooperation on industrial use-cases, tools (software, etc.) and methodology initiatives.

CONFORMITY ASSESSMENT

New conformity assessment schemes (grey stack), either for self-conformity assessment or third-party assessment, are required to prove systems conformity to risk, functional and trustworthiness requirements. Indeed, AI in critical systems will require defining of the new risks related to these

systems and adapting established testing and conformity assessment approaches to include AI capabilities and functions. This will enable significant performance improvement while ensuring an unprecedented level of safety maintained. This approach is key to gaining user trust and promoting the wide deployment of AI in strategic industries. For example, it is expected to develop new approaches for evolutive or even dynamic assessment of the system, and that tests and certification schemes will rely more and more on simulation or virtual testing to perform conformity assessment (in addition to field-testing) of AI systems. These approaches will be fundamental for safety demonstration.

France and Germany have already launched projects on these topics, such as confiance.ai, *Prissma (both* within the "Grand Défi" in France) *or KI-Absicherung* for autonomous driving. "Testing and experimentation Facilities" (TEF) promoted by the European Commission should be investigated as an opportunity for new cooperation at the European level.

STANDARDISATION AND REGULATORY SANDBOXES

In addition, this initiative will actively **promote standards that are in line with European values** and its socio-economic interest and digital sovereignty. France and Germany have ongoing cooperation involving Afnor (within the "Grand Défi" in France) and the German Standardisation roadmap on AI (for Germany). But also with CEN/CENLEC and other organisation bodies (ISO, IEEE, IEC etc.) standardisation initiatives for AI are ongoing.

In conclusion, following the signature of the Aachen Treaty in January 2019, France and Germany have signed in October 2019 a roadmap to strengthen the links between research and industrial actors of both countries. This roadmap aims at pooling German and French industrial capacities and the excellence of their innovation ecosystems to propose concrete bilateral projects on AI. Such projects must target the co-development of key technological building blocks for the competitiveness of both countries' economic sectors and the European industrial policy as a whole.

In this context, France and Germany are initiating discussions on the organisation and governance of joint Industrial & Trustworthy AI initiatives that will promote a common standardisation and technical framework - roadmap, as well as scientific and technical cooperation based on existing national initiatives (to produce results and benefit from existing innovation ecosystems quickly). At the same time, the representatives of French and German industries and research organisations shall pursue their dialogues and build up a comprehensive vision of how to develop and implement a European

Industrial AI ambition. However, this should not be seen as a bilateral initiative. It should be seen as a nucleus or blueprint in and for Europe!

APPENDIX: DESCRIPTION OF INDUSTRIAL & TRUSTWORTHY AI CHALLENGES

High-level System Specification

Encompassing all system specifications (operational design domains, risk and functional specifications, trustworthiness).

We are facing with AI a big challenge for the system and function requirements engineering discipline. Indeed, there is still no widely used and specifically tailored process to effectively and efficiently specifying a solution that uses machine learning. A key challenge is introducing increasingly formal expression of properties, specifications and requirements, for both risks, functional or trustworthiness specifications that can be mapped, measured and refined on machine learning-based designs. Hence, the development of essential specifications standards and a framework of basic requirements for, e.g. sectors would be necessary as a first step.

Conformity Assessment Schemes

Aligning the capabilities of AI-based application and systems with all the requirements

AI for high-risk applications will require adapting established new testing, and conformity assessment approaches, definition of adapted risk referential and new assessing methodologies better adapted to highly evolutive systems (updates, environment, use context/objectives…), developing new technical capabilities (especially with simulation or virtual testing) to enable improvement of performance obtained with AI while ensuring an unprecedented level of safety, security and trustworthiness.

Industrial & Trustworthy AI Engineering

Bringing ring forward all methodologies and tools to address trustworthy, safety and security AI requirements in the context of operational design domain, risk and functional requirements

- ➔ **Machine Learning Operations (MLOps)**: It aims to develop and maintain production machine learning seamless and efficient. The data science community generally agrees that it is an umbrella term for best practices and guiding principles around machine learning – not a single technical solution. MLOps applies to the entire lifecycle, from DataOps, model generation, software development to continuous integration/continuous delivery, orchestration, and deployment to enable diagnostics, governance, and business metrics supervision.

- **Data, Knowledge and Quality of data set:** Machine Learning, as the name implies, learns either from data or from experience (contrary to expert systems, which rely on handcrafted rules designed by human experts). When the machine is learning from data, either limited data or big data, the designer has to ensure the **quality of the knowledge** extracted from the data. This implies, for instance, the detection and also the control of bias. For example, learning to recognise road signs from daylight pictures may be suboptimal when the system is used at night. Moreover, the machine-learning algorithm might need to warn the user when incoming signals (e.g., images) to be classified are out of the distribution of the training set. For instance, a classifier for road signs should learn to tell the user that it couldn't conclude anything when confronted with a picture of a cat. These issues are encountered in both in offline supervised and unsupervised learning and in reinforcement learning. To further develop AI technologies, large volumes of cross-sectoral, unbiased, high-quality and trustworthy data and knowledge need to be made available. Data spaces, platforms and marketplaces are enablers, the key to unleashing the potential of such data and knowledge. There are however, important business, organisational and legal constraints that can block this scenario, such as the lack of motivation to share data due to ownership concerns (lack of trust; lack of foresight in not understanding the value of data or its sharing potential; lack of data valuation standards in marketplaces; legal blocks to the free-flow of data and the uncertainty around data policies). Additionally, significant technical challenges such as semantic interoperability, data verification and provenance support, quality and accuracy, decentralised data sharing and processing architectures, and maturity and uptake of privacy-preserving technologies for big data directly impact the data made available for sharing.
- **Embedded & Frugal AI:** The energy required to power cutting-edge AI has doubled roughly every 3.4 months—increasing 300,000 times between 2012 and 2018. That is faster than the rate at which computing power historically increased, the phenomenon known as Moore's Law. Dedicated embedded architecture (TPU, Neuromorphic, …), new approaches such as DataCentric (90% of energy consumption in moving data) will help, in conjunction with better training method to use fewer data, reuse already train model (transfer learning) to save energy and sustain the environment. In addition, industrial plants produce and consume large amounts of sensitive data that can't be easily transferred to the cloud due to the amount, its business sensitivity and legal restrictions. On the other hand, AIs are preferentially developed and trained in the cloud because the dedicated AI processors and data pools are available. To overcome this, technical approaches to learning across decentralised edge devices, such as federated learning, will be needed.
- **Explainable AI;** For gaining trust and confidence for any critical applications (where "critical" needs to be defined with clarity), one should be able to explain how AI applications came to a

specific result. Explainable AI (XAI) refers to those AI techniques aimed at explaining, to a given audience (AI engineers, end-users, and auditors), the details or reasons a model produces its output. Explainability will ensure the commitment of industrial users to measurable ethical, safety or robustness values and principles when using AI. Therefore, XAI targets bridging the gap between the complexity of the model to be explained and the cognitive skills of the audience for which explainability is sought. XAI should provide transparency about input data and the "rationale" behind the algorithm usage leading to the specific output. The algorithm itself need not necessarily be revealed in this case. Moreover, a step beyond XAI toward trustworthy-AI is Responsible AI, which denotes a set of principles to be met when deploying AI-based systems in practical scenarios: Fairness, Explainability, Human-Centric, Privacy Awareness, Accountability, Safety and Security.

➔ **Robust, Reliable & Safe AI**, from data collection to operation monitoring: Robustness ensures that an AI system maintains its level of performance under any circumstance, including unexpected interferences or environmental conditions. Global robustness is the ability of the system to perform the intended function in the presence of abnormal or unknown inputs and show that there is no side effects during operation and not included into specifications. Local robustness is the extent to which the system provides equivalent responses for similar inputs. Safe AI ensures that any action accomplished by the AI system is monitored to ensure that they are in a tolerable range or bring the system into a defined state in case problematic behaviour is detected during its operation. Many AI algorithms have shown that they can be fooled by data manipulation. Meanwhile, adversarial AIs have already been developed that are specialised in exploiting this weakness. There are already signs that attackers are beginning to develop their attacks with AIs so complex and dynamic that current detection mechanisms fail.

➔ **System tests & Evaluation** (use of simulation approaches, etc.) in conformity assessment and certification schemes: Incorporating artificial intelligence (AI) leveraging statistical machine learning (ML) into complex systems poses numerous challenges to traditional test and evaluation (T&E) methods and tools. As AI handles varying decision levels, the underlying ML needs the confidence to ensure testable, repeatable, and auditable decisions. Additionally, we need to understand failure modes and failure mitigation techniques. We need AI assurance–certifying ML and AI algorithms function as intended and are vulnerability free, either intentionally or unintentionally designed.

AI-based Functions

Establish the technical means to realise the functionalities of the AI-based business applications and services.

- **Sensing and Perception:** Sensing and Perception technologies create the information needed for successful learning, decision-making, and interaction. They encompass methods to access, assess, convert and aggregate signals representing real-world parameters into processable and communicable data assets that embody perception.

- **Knowledge and learning**: Data is the critical input for AI value; data needs to be transformed to become usable. This requires a wide range of data processing technologies, covering the transformation, cleaning, storage, documentation, sharing, modelling, simulation, synthesising and extracting of insights of all types of data both that gathered through sensing and perception as well as data acquired by other means, for example, financial transaction data, or marketing data. Combining both data-driven and knowledge-based models will establish the basis to a) support the fully automated enactment and actuation of decision, establishing a significantly higher level of automation and reliability of processes, b) develop safe, secure, bias-free and reliable AI functionalities and c) create sustainable digital twins along the complete lifecycle (from product design to operation) that provides value to AI data integration.

- **Reasoning and Decision Making** are at the heart of Artificial Intelligence. This technology area addresses optimisation, search, planning, diagnosis and relies on methods to ensure robustness and trustworthiness. It uses a set of defined knowledge (or business rules) to derive and manipulate data. A particular feature is based on a separation between the knowledge, which can be represented by various approaches such as rule-based inference, constraint solving, or reasoning algorithm, which uses knowledge to build a conclusion. These critical technologies bring value to multiple scenarios, ranging from human decision making to decision support systems, mixed, collective or distributed decision making. Different methods for decision-making can be utilised in all these scenarios, and constraints and uncertainty are taken into account. The quality of decisions depends heavily on the quality of input data and knowledge, including symbolic and non-symbolic data.

- **Action and Interaction** functions embody every aspect of digital and physical AI working together. Interactions occur between machines and objects, between machines, between people and machines and between environments and machines. Interactions are shaped by real-time data acquisition, stored information, long term knowledge accumulation and multiple modalities

and languages. There is often the need for regulatory compliance, especially when operating close to people.

AI innovation and trust demonstration and best practices

Establish the standardisation framework and innovation ecosystems to support European regulation adoption and business value creation in a secure environment

➔ **Experimentation and Sandboxing:** Experimentation is critical for AI-based applications and systems because of the need to deploy in the complex physical and digital environment while fulfilling high safety and reliability standards. Experimentation also plays a crucial role in innovation pipelines, being vital in supporting investment decisions. As the impact of regulation and conformity assessment on AI-based application development and deployment is highly complex -- especially when autonomous decision making or learning are involved -- a regulatory sandboxing environment enabling the high quality testing are needed.

➔ **Standardisation:** Standards and conformity assessments can be employed as a mechanism to leverage best practices and benchmarks to build trust in AI-based application and services. The increased collaboration between standardisation bodies, regulatory bodies and multidisciplinary teams of societal and industry stakeholders, including sectorial and citizen participation, should counter the fragmentation of standards. Further attention should be given to find innovative ways to simplify standardisation and conformity assessment-related processes and activities.

➔ **Cybersecurity:** AI requires strong Cybersecurity as a foundation. Cybersecurity is the conditio sine qua non for a resilient, trusted and reliable AI-based solutions. Only with security-by-design from the start and security-by-default in AI algorithms, this can be ensured. In addition, AI solutions need to be secured along the supply chain. This requires collaboration and co-creation across industries on a global level (like the Charter of Trust). Cybersecurity requirements for AI need to be developed and implemented and regularly tested.

ACKNOWLEDGEMENTS

Editors

Julien Chiaroni (Secretariat Général pour l'Investissement), Sonja Zillner (Siemens AG)

Contributors

Natalie Bertels (imec-KULeuven-CiTiP), Patrick Bezombes (ATDW), Yannick Bonhomme (IRT SystemX), Habiboulaye Amadou-Boubacar (Air Liquide), Loic Cantat (IRT SystemX), Gabriella Cattaneo (IDC), Lionel Cordesse (IRT Saint-Exupery), Edward Curry (Insight), Abdelkrim Doufene (IRT SystemX), Emmanuelle Escorihuela (Airbus), Ana García Robles (BDVA), Jon Ander Gómez (UPV), Thomas Hahn (Siemens AG), Stefan Jost-Dummer (Siemens AG), Frederic Jurie (Safran), Zoltan Mann (UDE), Juliette Mattioli (Thales), Andreas Metzger (Paluno/UDE), Yves Nicolas (Sopra Steria), Xavier Perrotton (Valeo), Milan Petkovic (Philips), Artur Romao (DECSIS), Simon Scerri (metaphactory), Maike Scholz (Telekom), Harald Schöning (Software AG), Marc Schoenauer (Inria), Kaoutar Sghiouer (Atos), Richard Stevens (IDC), Caj Södergard (VTT), Hubert Tardieu (Atos), François Terrier (CEA List), Eric Tordjman (Inria), Jean-Michel Tran (Naval Group), Henk-Jan Vink (TNO), Ray Walsh (DCU), Dimitris Zissis (MarineTraffic)

REFERENCES:

- Strategic Research, Innovation and Deployment Agenda, AI, Data and Robotics Partnership, September 2020 (joint initiative by BDVA, Claire, Ellis, EurAI, EuRobotics)
- Strategic road map of the French "Grand Défi" on trustworthy AI for industry, May 2019