IRT SystemX, a driving force for responsible AI to transform the industry of tomorrow

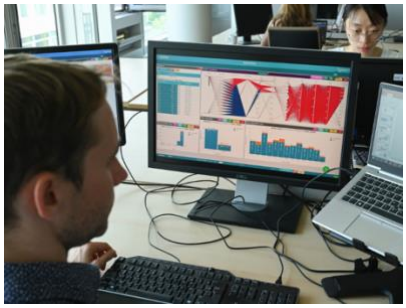
IRT SystemX

Centre d'intégration Nano-INNOV
Bât N3 – 2, boulevard Thomas Gobert
91120 Palaiseau

contact@irt-systemx.fr | www.irt-systemx.fr

Our vision

The Artificial Intelligence (AI) market is being driven by several key trends: the rise of generative AI and its application to numerous industrial and service-oriented use cases; the hybridization of traditional engineering tools with AI models; and the emergence of more sophisticated features made possible, in particular, by foundation models. These advances are expanding the functional scope of systems and raising expectations in terms of performance, versatility, and value creation. Industries are seeking to harness these opportunities to boost productivity and enhance the added value of their products, processes, and services.



However, they face two major challenges: transitioning from demonstrators to operational deployment, and adapting quickly to the constantly evolving scientific and regulatory state of the art.

We are convinced that the true value of AI lies not in the models themselves, but in their level of acceptability and their ability to be effectively integrated into systems. That is why **trust is at the heart of our work**—we believe it is the key to unlocking AI's full innovation potential for our industries.

For over ten years, we have been leading **multi-partner, cross-sector R&T projects**—such as **Confiance.ai**, a flagship program of the national AI strategy—to deliver tailored, by-design solutions to our partners.

An approach focused on industrial AI needs

Responsible, industrial AI is at the core of IRT SystemX's positioning. For AI to gain traction in strategic industrial sectors such as aerospace, automotive, energy, and defense, it must offer strong guarantees in terms of safety, security, and ethics. This is especially true for so-called **critical systems**, where there is no room for error.

Integrating AI into industrial systems is a major challenge. On the one hand, it requires an **end-to-end approach** to specify, design, assess, and operate these systems reliably throughout their entire lifecycle. On the other hand, companies must ensure that their AI-enabled systems **comply with European regulations**, including the **AI Act**.

To meet these needs, we ensure a **high level of trust** in the systems we develop, integrating principles of **robustness, explainability**, and **uncertainty quantification**. As part of the Confiance.ai program, SystemX, in collaboration with nearly **50 leading partners**, has established the state of the art in **Trustworthy AI engineering**: a **proven end-to-end methodology**, validated through industrial use cases, along with a **catalog of trust-enabling tools** available to scientific and industrial communities. These resources foster the secure, regulation-compliant design of industrial applications that incorporate trustworthy AI components.

Through this program, **IRT SystemX has become a key player in Trustworthy AI** within the French and European ecosystems. Thanks to its expertise and network, the institute now stands as a **true innovation and transformation driver**, promoting the emergence of responsible industrial AI. It has also played a critical role in the **development of European standards and norms** for evaluating and certifying AI components.

2030 strategy: deepen, sustain, industrialize

To ensure the long-term impact of the Confiance.ai program, **IRT SystemX is implementing a three-pronged strategy**:

- **Deepen** the results achieved by tackling challenges linked to new forms of AI, such as **Large Language Models (LLMs)**, generative AI, hybrid approaches combining AI with physical or human knowledge, and cybersecurity concerns through new R&D projects.
- **Sustain** and disseminate the methods and tools developed by launching the **European Trustworthy AI Association**, which aims to become a **driving force for the EU's industrial and responsible AI strategy**.
- **Industrialize** Trustworthy AI by guiding stakeholders toward solutions delivered by a network of experts and service providers tailored to their needs—be it **training** (e.g., a specialized master's in Trustworthy AI), **compliance support**, or tools for deploying trustworthy AI in industry.

Open source as a catalyst for AI adoption

IRT SystemX's strategy includes a commitment to **open source**, aimed at maximizing the impact of its work and encouraging the widest possible adoption—across Europe and globally—of its end-to-end methodology and tool catalog. This open approach enhances transparency, fosters collaboration, and enables a wide range of stakeholders to test, adopt, and improve the solutions.

By sharing its technological developments, SystemX aspires to become a central player in the **dissemination of best practices and standards for Trustworthy AI**, while also maintaining and expanding a strong **national and international community of excellence**.

Toward international engagement

Beyond establishing **scientific and technological collaborations** with international stakeholders, the institute is actively involved in **standardization committees** (e.g., ISO, CEN CENELEC, and AFNOR), as well as the **AI Trust Alliance**, to contribute to the development of **global AI standards**. Through these efforts, the institute reaffirms its commitment to playing a **leading role in the global digital transformation of industry**.

Our landmark impacts

Improving data reliability in critical industrial systems

IRT SystemX, in collaboration with Air Liquide, CEA, Inria, IRT Saint-Exupéry, and Thales, has developed TADkit: an open-source demonstrator designed to enhance data quality and identify biases in critical systems. It includes and enables the simultaneous use of more than ten innovative methods focused on the challenge of anomaly detection in time series data. This work was carried out as part of the Confiance.ai program led by IRT SystemX.

TADkit promises to transform the way industries analyze time series data. This demonstrator contributes to building monitoring tools for anomaly detection in industrial contexts and supports informed decision-making through artificial intelligence (AI) solutions.

This demonstrator was developed as part of the **Confiance.ai** R&D program to assess the quality and representativeness of data while identifying potential biases that could lead to faulty decisions in critical systems. The work conducted by IRT SystemX brings together more than ten original methods—combining deep neural network design, topological data analysis, and uncertainty quantification—focused on anomaly detection to provide insights into the operation of complex systems throughout their lifecycle.

TADkit consolidates these state-of-the-art AI methods into a single toolbox. This open-source demonstrator allows users to configure, compare, and combine advanced anomaly detection techniques to optimize their models and actively detect biases in data.

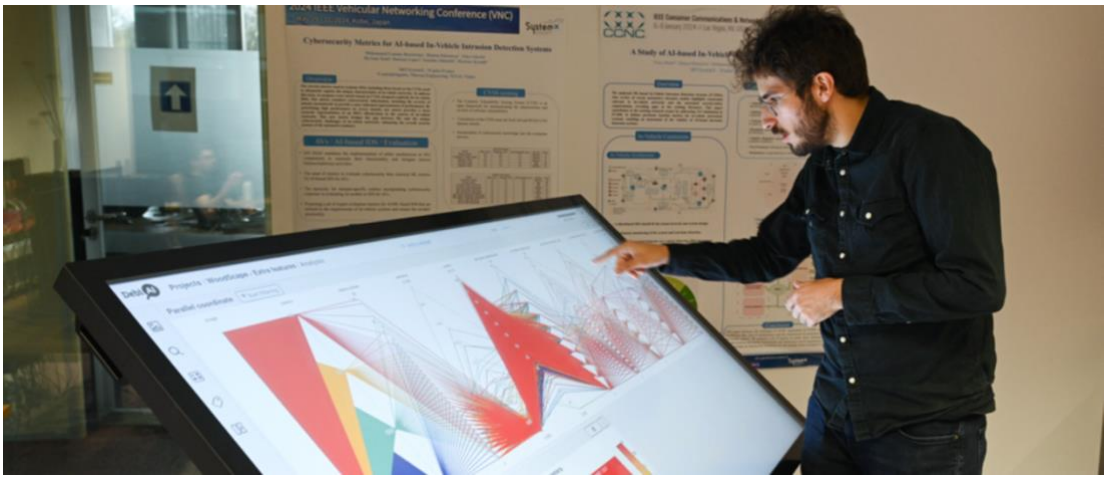
« TADkit brings together more than ten innovative, operational components that have been applied to use cases from Air Liquide and Naval Group as part of the Confiance.ai program. It has notably been the subject of a publication accepted in the prestigious Journal of Machine Learning Research (JMLR)*. Technically, it's a real Swiss Army knife—both user-friendly and already at a technology readiness level of 3. These features are essential for integrating cutting-edge AI solutions. »*

Martin Royer, Architect for the Data science and AI team, IRT SystemX

Want to learn more? Write to martin.royer@irt-systemx.fr

Optimizing data profiling in artificial intelligence projects

As part of the Confiance.ai program, IRT SystemX has developed DebiAI, an open-source data profiling tool designed to enhance dataset quality and improve the performance of artificial intelligence (AI) models.



The quality of datasets and the performance of models are critical to the success of R&D and industrial projects that rely on AI technologies. DebiAI addresses the growing needs of industrial stakeholders for effective use and control of AI models by offering a comprehensive solution for data analysis and instant bias detection. The tool enables data scientists to improve the accuracy and efficiency of their models, facilitating more informed decision-making based on trusted datasets.

DebiAI has established itself as a benchmark tool for exploring complex datasets and evaluating models in context. Tested on several use cases—including those led by **Thales** and **Safran**—DebiAI has proven its effectiveness in demanding industrial environments. Its strength lies in its open-source nature, which makes it easy to adopt and integrate. A public version of the tool is available for demonstration and hands-on use.

« DebiAI is more than just a data analysis tool. It's a performance catalyst for AI projects, part of the new family of HCAI (Human-Centered Artificial Intelligence) tools. It helps explore complex datasets and optimize models for reliable and robust results, all in a user-friendly way. »

Raphaël Braud, Head for the Data Science and AI team, IRT SystemX

Want to learn more? Write to raphael.braud@irt-systemx.fr, loic.cantat@irt-systemx.fr and remi.boyer@irt-systemx.fr

Boosting AI model performance in adverse conditions

As part of its Confiance.ai program, IRT SystemX has developed an open-source library called NeuralDE (Neural Domain Extension). This tool enables machine learning models—originally trained under ideal weather conditions—to maintain their performance in degraded environments. The work has significantly improved the robustness and operational reliability of Air Liquide's object-counting models in adverse weather conditions.

Machine learning (ML) models require vast amounts of data to train effectively. Ideally, this training data should cover all possible scenarios. However, unpredictable situations inevitably arise in industrial contexts, due to human and environmental factors. Additionally, some data is difficult to collect or exploit. The challenge lies in building ML models that can maintain performance even in scenarios where little or no real data is available.

Air Liquide uses AI-powered camera systems to automatically count outdoor inventory. Trained primarily during sunny daytime conditions, the counting tool's performance degrades noticeably at night or during bad weather (e.g., rain and darkness can reduce image resolution).

Several key challenges had to be addressed:

- How to detect these new scenarios in real time?
- How to conduct a sensitivity analysis of the system under such disruptions?
- How to ensure the system operates reliably in these conditions, while collecting, annotating, and augmenting new data for future training?

IRT SystemX developed a data pre-processing library for real-time deployment to eliminate visual noise (e.g., raindrops, snowflakes). **NeuralDE** successfully halved the nighttime counting error, achieving over **98% accuracy**, thanks to effective data pre-processing and better management of newly encountered scenarios.

« *We are proud of the work we've done, which has led to the deployment of NeuralDE in Air Liquide's production pipeline »,* **Martin Gonzalez**, Project Lead for Robustness, Uncertainty, and Monitoring, Confiance.ai / IRT SystemX

Want to learn more? Write to martin.gonzalez@irt-systemx.fr

Revolutionizing the supervision and control of critical systems

InteractiveAI, a platform born from a collaboration between IRT SystemX, Dassault Aviation, Orange, RTE, and SNCF, supports the supervision and control of critical and complex systems through an innovative approach combining interactive AI technologies with human-machine interaction.

InteractiveAI offers an interactive assistant designed to support operators in their decision-making processes. What sets it apart is its ability to incorporate human factors, enabling more effective use of the assistant and optimizing AI performance.

Built on user-centered design principles, the platform provides an **open-source framework** for industries managing critical systems (such as aircraft, automobiles, and trains) or supervising sensitive networks (including transport, energy, and telecommunications). It enables them to customize assistant interfaces and easily integrate advanced AI modules—delivering tailored solutions for complex, high-stakes environments.

*« Developed as part of IRT SystemX's **CAB project (Cockpit and Bidirectional Assistant)**, the InteractiveAI platform has led to the creation of bidirectional assistants tailored to the specific needs of our partners. These assistants support operators in carrying out daily tasks in complex environments while paving the way for industrial-scale deployment of interactive assistants. »*

Maroua Meddeb, Project manager, IRT SystemX

Focus: concrete results with European and international reach

4 technology transfers to CAB project partners (RTE, Orange) and European players (Flatlandet, EnliteAI)

- InteractiveAI serves as the technological foundation for the H2020 AI4REALNET project, strengthening its position as a reference platform in the field of interactive AI
- The project has led to the publication of 8 scientific papers, contributing to the advancement of knowledge in AI and human-machine interaction

Want to read more? Write to maroua.meddeb@irt-systemx.fr

Training the future experts of trustworthy AI

Led by IRT SystemX and CentraleSupélec Exed, the Trustworthy Artificial Intelligence Advanced Master's Program (Mastère Spécialisé® Intelligence Artificielle de Confiance) aims to train a new generation of professionals in the design, development, and deployment of trustworthy AI solutions for industry and services. This program was co-designed in close collaboration with CentraleSupélec and industrial partners, based on their expressed needs and the technological challenges identified within the Confiance.ai initiative.

The partnership between IRT SystemX and CentraleSupélec Exed combines the academic excellence of CentraleSupélec with the applied expertise of IRT SystemX in developing reliable and innovative AI solutions. This prestigious, certified program has been recognized by the Conférence des Grandes Écoles, which adds credibility and momentum to the initiative. The training combines cutting-edge theoretical knowledge with practical applications, with a strong focus on trust and ethics in AI systems.

The first cohort began in September 2024, bringing together 12 young graduates and professionals from across the fields of software, hardware, and systems engineering—many of whom are actively working on the integration of AI into industrial systems. Throughout the program, they will explore key topics such as AI ethics, transparency and security of AI-integrated systems, as well as the reliability, explainability, and robustness of algorithms. The knowledge shared by research engineers from the institute and the skills developed by the participants will be in high demand in the coming years.

« After permeating our personal lives and leisure, artificial intelligence is now disrupting the business world—including companies working with systems that have high demands in terms of safety, reliability, and performance, and therefore require trustworthy AI adapted to their use cases. Integrating AI into so-called critical systems is not just a technological challenge. Two additional hurdles must be addressed. First is the ability to incorporate trustworthy AI into a systems engineering approach—from design through to testing and validation. This requires new methodologies and tools, which are being developed in programs like Confiance.ai, of which we are a partner. The second challenge is skills and human resources. These must also be adapted to the specific needs of AI for critical systems. The specialized Master's in trustworthy AI engineering, co-developed by CentraleSupélec and IRT SystemX, responds to this need. All the stakeholders of Confiance.ai, as well as those behind the AI Manifesto, and more broadly the industrial sector, welcome this initiative—it meets an urgent demand. »,

Marko Erman, SVP Chief Scientific Officer, Thales

« Artificial intelligence combines applied mathematics, statistics, and computer science. It has existed for over 70 years and has made remarkable progress. AI has outperformed humans in games such as chess and Go, and has helped doctors by automatically analyzing medical images, enabling faster and more reliable diagnoses. It has also driven the creation of powerful language models and chatbots thanks to natural language processing. However, the most impressive and useful breakthroughs are yet to come. A major challenge is moving from proof-of-concept to industrial-grade, trustworthy products.»,

Hugues Talbot, Professor, Centre for Visual Computing,
CentraleSupélec, Inria, Université Paris-Saclay

Want to learn more? claudiu.balan@irt-systemx.fr

DQM: an open source library to revolutionize data quality evaluation for industrial AI

As part of the Confiance.ai program, IRT SystemX, in collaboration with Atos and CEA, has developed DQM (Data Quality Metrics), an open-source library designed in Python. This tool enables the evaluation of data quality used in the development and assessment of artificial intelligence (AI) models—particularly within complex industrial environments..

Data quality is essential to ensure the reliability of AI models. The DQM library provides a concrete response to this challenge by offering relevant and interpretable quality attributes that assess critical aspects such as the representativeness and coverage of data within specific operational domains.

The institute's teams developed two main categories of metrics:

1. Data-intrinsic metrics (e.g., representativeness, diversity), to evaluate dataset quality before an AI model is implemented;
2. System-dependent metrics (e.g., data-model coverage), which measure the impact of data on system performance once integrated into a model.

DQM was designed as a standalone Python package, making it easy to use independently or to integrate into other tools, such as DebiAI. The library has already been integrated into the end-to-end methodologies of major industrial players such as Naval Group and Valeo, strengthening their ability to accurately assess data quality.

« The potential of the DQM library is very promising. Our experiments have shown its effectiveness in providing a deep understanding of the data used in machine learning workflows. Integrated into the European Trustworthy Foundation created by the Confiance.ai community, the library is generating strong interest and is paving the way for new applications across various industrial sectors. Furthermore, a scientific paper detailing the library's contributions was published in ATRACC. »

Faouzi Adjed, research engineer and Data AI architect, IRT SystemX

Want to learn more? Write to faouzi.adjed@irt-systemx.fr

Contact us!

Interested in collaborating with us? Want to learn more about our R&D activities?

Contact **Raphaël Braud**, Head of the Data Science and AI team, IRT SystemX:
raphael.braud@irt-systemx.fr

More information on www.irt-systemx.fr/en