POSITIVE AI

Research Theme – IRT SystemX Projects from a Technological Research Institute

Summary of the Webinar - July 2025

Nicolas Rebierre
Head of Research and Development, IRT SystemX

Preamble

This summary was generated from the text transcription of the Webinar using ChatGPT 4, formatted by the Positive AI team and validated by the host.

Introduction
The webinar presented IRT SystemX's work on "trustworthy AI" engineering and the post-Confiance.ai roadmap: methods, tools, and a new European open-source association to industrialize and scale results. The talk covered the "W-model" for AI systems engineering, concrete tooling (robustness, explainability, uncertainty), governance with industry leads, education plans, and links with standards/regulation.

Main Points Discussed

1) Why trustworthy AI engineering matters

- High-stakes use cases (automotive, aerospace, manufacturing inspection, etc.) make the “cost of error” high; AI’s black-box aspects require updated engineering methods.
- The goal is to control risk and quality at system level, not just at model level, so that products can be sold and adopted with confidence.

2) Confiance.ai program: scope and outputs

- Multi-year collaborative R&D (2022–2024) with ~50 partners (large French industrials, research bodies, startups), focused on “classical” AI (CV, time series, tabular, limited NLP).
- Delivered ~180 assets: ~130 methodological guides + software tools gathered in a public catalogue.
- Outcomes also include scientific publications and contributions to standardization (CEN/CENELEC, AFNOR, ISO/IEC), with ongoing work toward harmonized standards.

3) End-to-end “W-model” for AI systems

- Extends the classic V-model to address AI components and system-level trust: from operational design domain (ODD) specification → architecture → component implementation & verification → system verification & validation.
- Emphasizes documentation, roles, artifacts, and verification activities specific to AI and “components including the model plus surrounding software/hardware safeguards.”

4) Tooling highlights (from the catalogue)

- **MOOD**: robustness analysis (e.g., how classification accuracy evolves under perturbations like blur, lighting).
- **Tatkit**: anomaly detection toolbox for time-series; side-by-side algorithm evaluation.
- **Heatmap/attribution methods**: explainability for images and time-series (what regions/signals drove a decision).
- **Conformal prediction (“PUNK”)**: predictive intervals/uncertainty bounds for regression and object detection (e.g., probability-guaranteed bounding boxes).

5) From R&D to industrialization: a European open-source association

- A non-profit, industry-oriented association is being set up to **harden, maintain, and support** the open assets (quality gates, test plans, IP/licensing, maintenance/support).

- Two-tier strategy: (1) open-source core maintained with industrial priorities; (2) an ecosystem of service/solution providers (deployment, compliance, training, integrated platforms).
- Rationale for open source: accessibility, transparency (trust), autonomy/sovereignty, cost-sharing, and a proven model for global collaboration.

6) Ecosystem & activities

- **Working groups**: Industry (needs/priorities), Science (roadmap & seminars), Standardization (push/pull with norms), Communication (events), Education (programs & MOOCs).
- **Events**: monthly member meetups, open scientific seminars, and an annual "Trustworthy AI Summit."
- **Education**: executive Master with CentraleSupélec/IRT SystemX; integration of content in initial curricula with other schools; MOOC development underway.

7) Membership model & governance

- Three tiers: **Engage** (discover events + catalogue/body of knowledge), **Use** (tool support + WG participation), **Lead** (co-fund portfolio, steer priorities, governance seats).
- An initial group of industrial "leads" (from the Confiance.ai cohort) is funding the ramp-up, with an open design to welcome additional leads.

8) Standards & regulation interface

- Continuous monitoring of EU/ISO work; contributions to definitions and evaluation approaches for high-risk AI.
- Goal: align methods/tools with evolving regulation (including future harmonized standards) and make industrial compliance feasible.

Q&A Highlights & Strategic Debate

- **Agile vs V/W**: Not contradictory—apply W-phases per increment/epic with lighter artifacts; iterate frequently while preserving specification/verification discipline.
- **Beyond safety-critical**: Tools/methods are applicable to broader "at-risk" and even moderate-risk business use cases (contracts analysis, market monitoring, etc.); rigor scales with stakes.
- **Adoption hurdle**: Outside directly regulated/high-risk sectors, executive buy-in and change management are often the main blockers; governance and business-level messaging matter.
- **Education pipeline**: Executive programs exist; initial-curriculum integration is expanding with French and European schools.

- **International cooperation & sovereignty**: Open to collaboration (US, Canada, Japan, Australia, Singapore), with constraints where export-control or autonomy would be jeopardized.

Conclusion

Trustworthy AI requires system-level engineering, documented processes, and proven tools across the lifecycle. Confiance.ai created a robust starting corpus (methods, tools, publications). The new European open-source association aims to industrialize, maintain, and support these assets, while interfacing with standards and building talent pipelines. For organizations, combining agile delivery with W-model guardrails, inventorying AI systems, documenting/monitoring components, and engaging leadership and education are key to scaling AI with confidence