Trustworthy Autonomous Wheelchairs in Practice: Research Highlights from REXASI-PRO

Francesco Flammini - IDSIA USI-SUPSI, Lugano, Switzerland; University of Florence, Italy

francesco.flammini@supsi.ch

How distributed sensing, explainable AI, runtime assurance, and human-centered evaluation were combined in a European research programme on autonomous assistive mobility.

Why assistive autonomy is a systems problem

Autonomous wheelchairs place artificial intelligence in immediate physical and social contact with people. They must navigate crowded indoor spaces, interpret incomplete and sometimes contradictory sensor data, preserve the user's ability to intervene, and behave in ways that nearby pedestrians can anticipate. For the IEEE Systems, Man, and Cybernetics community, this is a particularly revealing application: safe performance emerges from the interaction of engineered components, human decision makers, communication networks, and the environment rather than from one algorithm in isolation.

REXASI-PRO - REliable and eXplainable Swarm Intelligence for People with Reduced mObility - was a Horizon Europe project that ran from 2022 to 2025. It investigated how an autonomous wheelchair could benefit from aerial robots, wall-mounted cameras, and a central orchestrator while retaining local safety functions and meaningful human control. The programme addressed trustworthy sensing, social navigation, explainability, formal verification, runtime monitoring, cybersecurity, ethics, and resource-aware AI. This article highlights representative research results and the integrated trials that connected them.

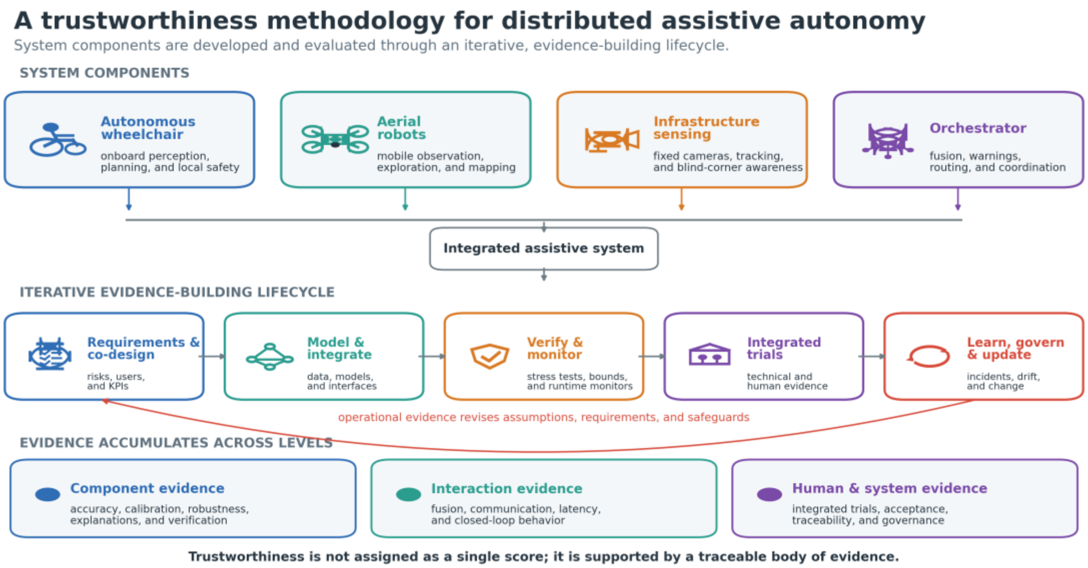

Figure 1. REXASI-PRO as a distributed assistive system and an evidence-building lifecycle. Wheelchairs, aerial robots, infrastructure sensing, and orchestration are developed through requirements and co-design, integration, assurance, trials, and operational learning.

The research system

The project organized the technology around four interacting components. The autonomous wheelchair provided onboard perception, local planning, control, and an independent safety layer. Aerial robots extended observation and mapping beyond the wheelchair's field of view. Fixed cameras supported people tracking, crowd awareness, and blind-corner monitoring. The orchestrator coordinated information exchange, warnings, and route recommendations. This separation was intentional: the wheelchair had to remain locally safe even when global information was delayed or unavailable, while system-level intelligence could improve anticipation and planning when reliable external observations were present.

The engineering process followed an iterative logic rather than a linear sequence. Requirements for safety, security, ethics, explainability, and user control were connected to measurable tests; components were evaluated in isolation and then incrementally integrated; runtime monitors were added where open-world conditions could invalidate design assumptions; and evidence from trials was used to revise models and safeguards. Figure 1 summarizes this relationship between the technical system and the lifecycle used to build confidence in it.

Representative research results

Trustable sensing and risk-aware road crossing

One research thread used a road-crossing scenario to make perception uncertainty operational. An onboard wheelchair camera, an aerial camera, and distance sensing observed an approaching vehicle. A continuous danger function combined estimated distance, speed, and acceleration to support a proceed-or-wait decision. The experiments compared single sensors with distance, danger, and voting fusion. Distance fusion achieved the strongest overall balance in the laboratory dataset, with 77.25% decision accuracy and an F1 score of 69.59%, while danger fusion produced the highest recall at 73.84%. More importantly, fusion smoothed unstable measurements and reduced dependence on any one sensor [2], [3].

The same problem was then modeled with Bayesian Networks and Dynamic Bayesian Networks (DBNs), which explicitly represented missing observations, uncertainty, and temporal dependencies. The dynamic models outperformed both the arithmetic baseline and the static Bayesian model on the evaluated runs. Performance improved as temporal context increased up to an intermediate number of time slices, after which model variance, data requirements, and inference cost became more significant [6]. This is a useful SMC lesson: trustworthy sensing is not only about maximizing nominal accuracy; it is about carrying uncertainty into the decision layer and understanding the tradeoff between model fidelity and real-time feasibility.

Explainability, verification, and runtime assurance

A second thread combined explainable AI with reliability guarantees. Simulation data were used to identify probabilistic and conformal safety regions for collision and deadlock avoidance. These regions associated navigation parameters with bounded error, while local rules near their boundaries helped engineers understand why a configuration was accepted or rejected. In project experiments, conformal safety regions kept the selected error rate within a 10% bound in the simulated setting [4]. Complementary work addressed out-of-distribution detection, so that unfamiliar operating data could trigger caution rather than being treated as ordinary inputs.

Formal methods and runtime monitoring answered different but complementary questions. Design-time verification examined whether neural agents satisfied temporal requirements under stated assumptions. Predictive monitoring used conformal inference to estimate, during operation, whether requirements such as remaining collision-free over a future horizon were likely to remain satisfied [5]. The practical objective was not merely to produce another confidence score, but to support mitigation: slow down, stop, reroute, or ask for assistance when confidence deteriorated.

Human-in-the-loop navigation, orchestration, and efficient deployment

The project also developed the Navground and Navground-Learning toolchain for repeatable multi-agent navigation experiments and learning-based policies. Virtual and mixed-reality testbeds allowed humans to interact with simulated robots in narrow passages and collaborative navigation scenarios before equivalent behaviors were exercised on physical platforms. A voice-based human-machine interface was integrated on the wheelchair, reflecting the principle that automation should support rather than displace user agency.

At system level, the orchestrator received information from fixed and mobile sensing, estimated crowd conditions, issued collision warnings, and supported online routing decisions. Secure MQTT-based communication and device authorization were part of this architecture. Resource awareness was treated as part of trustworthiness: dataset reduction, model compression, and topology-based analysis were investigated to reduce the training and runtime cost of AI components [7]. This matters because monitoring, redundancy, secure communication, and explanation all compete for limited edge-computing resources.

From methods to integrated trials

The final evaluation campaign in Genoa in June 2025 connected these research threads in a shared indoor environment. In one test, infrastructure cameras detected a running person outside the wheelchair's onboard field of view. The predicted path intersection was communicated so that an emergency stop could be executed. Other tests demonstrated crowd detection using fixed cameras and aerial robots, projection of crowd-density estimates onto a navigation graph, and online rerouting of an autonomously operating wheelchair around a detected crowd.

The campaign also included user-facing evaluation. Sixteen participants operated the wheelchairs and 36 participants acted as nearby pedestrians, for a total of 52 people. User Experience Questionnaire and Godspeed categories were used alongside technical analyses of navigation behavior. These trials do not constitute a certification case, but they provide a stronger form of evidence than isolated model benchmarks: perception, communication, orchestration, local safety, autonomous navigation, and human acceptance were observed together.

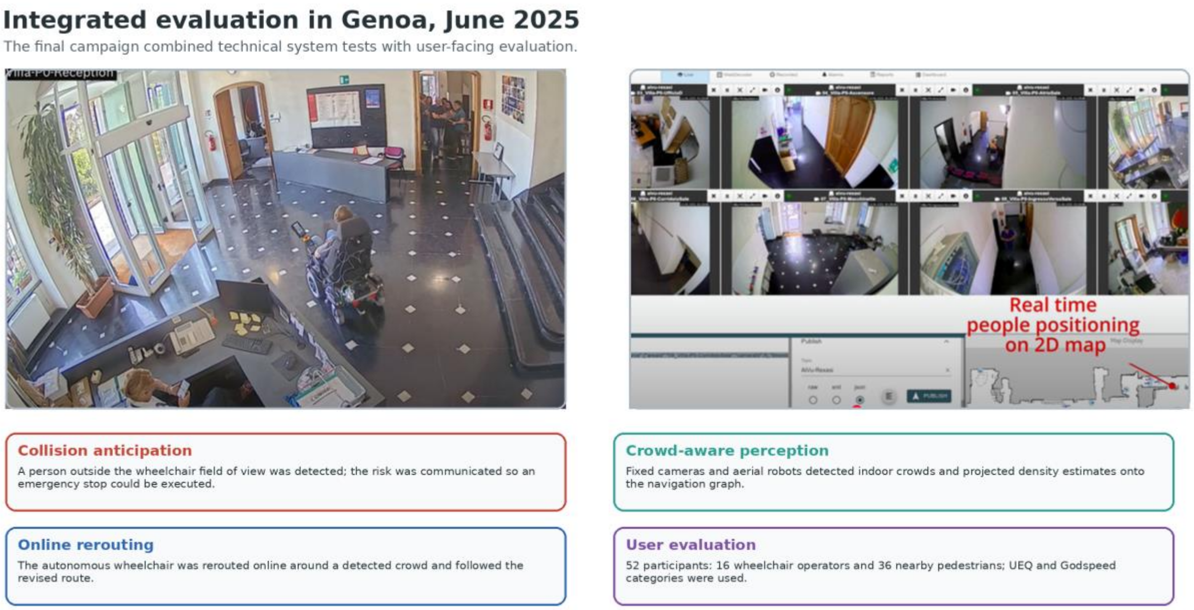

Figure 2. Integrated evaluation in Genoa, June 2025. The campaign combined autonomous wheelchair operation, infrastructure and aerial perception, collision anticipation, crowd-aware routing, and user-facing assessment.

Lessons for the SMC community

The project suggests four transferable lessons for human-machine systems and cyber-physical autonomy:

- **Trustworthiness is a system property.** A dependable model cannot compensate for weak interfaces, unsafe fallback behavior, unreliable communication, or poor human oversight.
- **Uncertainty must reach the action layer.** Probabilistic perception and calibrated monitoring are valuable when they change what the system does under ambiguity.
- **Design-time verification and runtime monitoring are complementary.** One establishes guarantees under explicit assumptions; the other checks whether those assumptions remain credible during operation.
- **Human and integration evidence are indispensable.** A system intended for shared environments must be judged by technical behavior, user control, bystander interaction, and organizational readiness together.

REXASI-PRO did not solve every problem associated with trustworthy assistive autonomy. Long-term deployment, heterogeneous users and environments, compositional certification, and maintenance after software updates remain open. Its main contribution is therefore not a claim of completed trustworthiness, but a set of methods, tools, and integrated demonstrations showing how evidence can be accumulated across components, interactions, and human use.

Further information

- [REXASI-PRO project website]
- [Project outcomes and publications]
- [REXASI-PRO Zenodo community]

Selected references

[1] F. Flammini et al., "Towards Trustworthy Autonomous Systems: Taxonomies and Future Perspectives," IEEE Transactions on Emerging Topics in Computing, vol. 12, no. 4, pp. 1057-1072, 2024.

[2] F. Corradini, F. Flammini, and A. Antonucci, "Probabilistic Modelling for Trustworthy Artificial Intelligence in Drone-Supported Autonomous Wheelchairs," Proc. First International Symposium on Trustworthy Autonomous Systems, ACM, 2023.

[3] C. Grigioni, F. Corradini, A. Antonucci, J. Guzzi, and F. Flammini, "Safe Road-Crossing by Autonomous Wheelchairs: A Novel Dataset and Its Evaluation," Computer Safety, Reliability, and Security Workshops, Springer, 2024.

[4] S. Narteni, A. Carlevaro, J. Guzzi, and M. Mongelli, "Ensuring Safe Social Navigation via Explainable Probabilistic and Conformal Safety Regions," Explainable Artificial Intelligence, CCIS 2156, pp. 396-417, Springer, 2024.

[5] F. Cairoli, N. Paoletti, and L. Bortolussi, "Conformal Quantitative Predictive Monitoring of STL Requirements for Stochastic Processes," Proc. 26th ACM International Conference on Hybrid Systems: Computation and Control, 2023.

[6] F. Corradini, C. Grigioni, A. Antonucci, J. Guzzi, and F. Flammini, "Balancing Accuracy and Interpretability in Multi-Sensor Fusion through Dynamic Bayesian Networks," Proc. TRUST-AI: The European Workshop on Trustworthy AI, CEUR Workshop Proceedings, vol. 4132, 2025.

[7] J. Perera-Lago et al., "An In-Depth Analysis of Data Reduction Methods for Sustainable Deep Learning," Open Research Europe, vol. 4, art. 101, 2024.

About the author

Francesco Flammini is Professor of Trustworthy Autonomous Systems at IDSIA USI-SUPSI and Full Professor at the University of Florence. He leads the Trustworthy Autonomous Systems group at IDSIA and works on trustworthy AI, safety and security of cyber-physical systems, formal methods, and resilience engineering. Within REXASI-PRO, he led the Swiss unit and the work on trustable sensing. He is a Senior Member of IEEE and serves on the IEEE SMC Society Board of Governors.

Contact: francesco.flammini@supsi.ch | tas.idsia.ch/people/francesco-flammini