

Final Report

Project information

Project Acronym	GALICIA
Coach Name	JUAN JUAN
EU Organization	NOVARECKON SRL
CA/US Organization	MIND IN A BOX

EU Organization Members

- Alberto Stefanini (Novareckon)¹
- Alberta Bertin (Novareckon)
- Andrea Ceccaroni (Novareckon)
- Elenora Dorissi (Novareckon)²
- Christian Violi (Novareckon)
- Lorenzo Vandoni (from Hal Service, end user company)

CA/US Organization Members

- Jeremie Farret (Mind in a Box)
- Francois Monette (Mind in a Box)

¹ Eleonora and Alberto edited the whole Report.

Project description

The GALICIA project (**Generative AI with Cybersecurity for Internet Applications development**) is a European research experiment funded under the **NGI Sargasso initiative**, part of the European Commission's Next Generation Internet (NGI) program. NGI Sargasso aimed to foster collaboration between European and US/Canadian innovators to develop next-generation internet solutions, focusing on trust, security, and resilience.

GALICIA focused on exploring the integration of **generative AI techniques** with **formal modeling** to support the trustworthy development of Internet-based applications, with specific regard to cyber security threats. The project investigated how recent advances in AI, particularly in automated code generation and system modeling, could be combined with formal verification, validation, and cybersecurity-by-design approaches to enhance the security, reliability, and quality of software applications.

GALICIA developed workflows and demonstrators that showed how formal methods, model-based engineering, and AI-generated artifacts could be combined to ensure more secure and trustworthy internet application development processes. Special attention was devoted to reducing the risks associated with software vulnerabilities, insecure protocols, and lack of assurance in AI-generated components.

The project contributed to bridging the gap between agile AI-driven development methods and the stringent cybersecurity requirements increasingly demanded for modern internet infrastructures.

Table of Contents

Description of activities carried out during the Program.....	4
Impact analysis and/or feasibility study	10
Publications	15



Description of activities carried out during the Program

Kindly adhere to the outlined scheme while formulating your report, adding any other relevant information:

Describe the Use Case: provide a detailed description of the use case your project addresses. Explain the context and relevance of your project within this use case.

The GALICIA project addresses [critical use cases](#) at the intersection of software reliability and AI-assisted development, enabling the generation and formal verification of source code from natural language requirements. As described in the sections below, these use cases concern ICT applications, formalizing mathematics, and network protocols. As the demand grows for trustworthy, explainable, and safe AI-generated code, GALICIA provides an innovative approach that integrates large language models (LLMs) with formal verification techniques to ensure the generated code is not only functional but also secure and logically consistent. This approach was verified across three core domains explored through real-world case studies:

1. **ICT Applications** – The “WiC Advisor” scenario involves managing customer data, verifying internet coverage, and generating product quotes through external API calls. GALICIA demonstrated the ability to generate CRUD operations and verify their behavior under real-life conditions such as malformed input or service unavailability. This shows the platform’s applicability to modern web and IoT systems that rely on service orchestration and real-time responsiveness.
2. **Formalizing Mathematics** – By translating Euclidean axioms and Fibonacci series into formal models and generating proofs in Coq, Prolog, and Python, GALICIA showcases how structured knowledge domains can be digitized and explored through AI. This has implications for education, mathematical tooling, and logic-driven development environments.
3. **Network Protocols** – The modeling and verification of the TCP three-way handshake demonstrate GALICIA’s potential in verifying communication reliability in safety-critical systems. It ensures correctness across message transmissions, handling retries and acknowledgments, which are fundamental to secure networking.

Overall, GALICIA is designed to bridge the gap between user-friendly natural language inputs and robust, formally verified code, paving the way for safer, AI-assisted code development in complex, real-world scenarios.

Integration of Social Sciences and Humanities (if applicable): *for projects under topics indicating the need for the integration of social sciences and humanities, explain the role of these disciplines in the project thus far. Detail how they have contributed to your project’s objectives and outcomes.*

Does not apply



Project Objectives: *list the specific objectives for the project detailed at the Individual Mentoring Plan. Provide a short summary of the progress made towards the achievement of each project objective. Highlight significant activities that have contributed to these achievements¹. Please ensure that you report on objectives that have not been fully achieved or are not on schedule as well.*

KPI 1 – [GALICIA Website](#)

Objective: *Create and publish a dedicated website.*

Status: Achieved

Progress: *Website launched in Month 3. It serves as the central platform for sharing updates, publications, and public access to the prototype.*

Activities: *Design, development, and integration of project communication materials.*

KPI 2 – [First Newsletter](#)

Objective: *Disseminate initial project outcomes and context, including call to action for participate at the GALICIA Survey.*

Status: Achieved

Progress: *Distributed in Month 3, including an overview of objectives and case study plans and a presentation of partners involved in the project.*

Activities: *Drafting, review, and stakeholder distribution via mailing list and website.*

KPI 3 – [Component Selection](#)

Objective: *Choose key technical components for platform development.*

Status: Achieved

Progress: *Completed in Month 5 Core LLMs (e.g., GPT-4, LLAMA), formal verification tools (Coq, Prolog), and APIs selected.*

Activities: *Internal review and expert consultation.*

KPI 4 – [Platform Logical Architecture & Requirements](#)

Objective: *Define platform structure and specs.*

Status: Achieved

¹ overviews of the interviews conducted and analyses of the two panels' replies were published at various stages during project development, namely in four documents cited in this section: *KPI7 – Terms of Reference of the mid-term Workshop* (section 3.2 Survey Results, p. 4 ff.); *KPI9 – Proceedings of the mid-term GALICIA workshop* (section 3.2 Survey Results, p. 4 ff.); *KPI10 – Platform Evaluation Results* (analysis of the first panel's replies in Chapter 4 Reply Analysis, p. 9 ff.); and *KPI13 – Platform Evaluation and Verification* (analysis of the second panel's replies in Chapter 4 Reply Analysis, p. 11 ff). This latter document also provides an overview of the platform usage so far in Chapter 2 Usage Statistics, p. 6 ff.

Progress: Logical architecture documented in Month 5, outlining how formal models, natural language inputs, and code generation interact.

Activities: UML diagramming and tech requirement mapping.



KPI 5 – [Platform Interface Mock-up](#)

Objective: Deliver a mock-up of the platform.

Status: Achieved

Progress: Interface prototypes completed and reviewed in Month 5 Used during user workshops for early feedback.

Activities: UI/UX wireframing and validation with stakeholders.

KPI 6 – [Platform Delivery](#)

Objective: Deploy GALICIA prototype.

Status: Achieved

Progress: Working version tested across three major use cases (IoT, geometry, and TCP/IP).

Activities: Full stack implementation, integration of model-checking logic.

KPI 7 – [Mid-Term Workshop ToR and Questionnaire](#)

Objective: Plan the mid-term workshop.

Status: Achieved

Progress: Workshop held in March 2025. Questionnaire evaluated usability and functional gaps.

Activities: Agenda preparation, SUS questionnaire development, expert invitation.

KPI 8 – [Second Newsletter](#)

Objective: Communicate mid-project results.

Status: Achieved

Progress: Published in April 2025. Summarized use cases and early feedback.

Activities: The newsletter presents a summary of the workshop, the test cases and the platform.

KPI 9 – [Workshop Report](#)

Objective: Document participation and outcomes.

Status: Achieved

Progress: Detailed participant analysis and SUS scoring compiled.

Activities: Data analysis, synthesis of replies, and recommendations.

KPI 10 – [Platform Evaluation Results](#)



Objective: Publish a formal usability evaluation report.

Status: Achieved

Progress: Completed April 2025. Highlights both strengths and improvement areas.

Activities: SUS-based user testing, expert feedback aggregation.

KPI 11 – [Third Newsletter](#)

Objective: Final dissemination of project insights.

Status: Achieved

Progress: verification insights and platform evolution.

Planned Activities: The Newsletter presents the final workshop results, the Platform Evaluation (KPI13), consideration on the Final Report and conclusion of the project.

KPI 12 – [Final Workshop ToR and Questionnaire](#)

Objective: Results of the final Workshop

Status: Achieved

Progress: held on the 15th of April in Montreal, during the WSAI

Planned Activities: Presentation of the GALICIA Platform, core objectives, technical architecture and preliminary results. These outreach efforts promoted trust in AI, validated its technical approach, and positioned the project for future collaboration and adoption.

KPI 13 – [Platform Verification and Evaluation](#)

Objective: Benchmark GALICIA-generated code vs direct LLM output.

Status: Achieved

Progress: Functional tests are ongoing; comparison metrics being developed.

Delays: Extra time required for reliable dataset construction and test coverage expansion.

Planned Activities: Complete testing and statistical analysis of success rates.

Work Carried Out: explain the work carried out during the program per Stage, providing details of the work conducted by each beneficiary or affiliated entity involved in the project. Describe the roles and contributions of each entity to the project's overall progress.

- Stage 1 - **Initial Research and Requirements Definition** - This stage laid the foundation for the GALICIA platform. The work was driven by a joint effort between NOVARECKON (EU lead) and Mind in a Box (CA/US partner).
 - **NOVARECKON** coordinated the requirement-gathering process, conducted risk assessments, and facilitated discussions around platform architecture. They were also responsible for drafting [KPI4](#), the public deliverable on logical architecture and requirements.

- **Mind in a Box** contributed technical insights into generative AI capabilities and limitations. They conducted initial benchmarking of LLM behavior and identified key failure points, such as output inconsistency and semantic ambiguity. This phase concluded with the publication of the architecture document (KPI4) and interface mockups (KPI5), ensuring alignment with end-user expectations.
- Stage 2 - **Methodology Design and Platform Development** - Once the requirements were established, the partners collaboratively designed the hybrid methodology and began implementing the platform.
 - **Mind in a Box** led the technical implementation of the LLAMA-based code generation module, working closely on the integration of AI prompt engineering into the platform.
 - **NOVARECKON** focused on integrating formal verification tools, particularly Event-B, into the pipeline. They also ensured user interface requirements were met, working with end users such as Hal Service to validate mockups and workflows. During this phase, the platform's four-stage methodology—prompting, transformation, verification, and feedback—was tested iteratively, resulting in a modular, human-supervised development workflow.
- Stage 3 - **Demonstrator Development and Testing** - This phase focused on validating the methodology and platform through practical use cases.
 - **NOVARECKON** coordinated the development of three demonstrators: TCP/IP protocol modeling, mathematical proof generation, and CRUD operations. They ensured alignment with real-world application domains and worked closely with Hal Service to simulate realistic user requirements for IoT scenarios.
 - **Mind in a Box** focused on AI prompt optimization and ensured coherence between AI-generated output and formal models. They also contributed to the creation of benchmarking criteria for the demonstrators, used later in KPI13.

The demonstrators were critical in testing the hybrid pipeline and revealing both productivity gains and model limitations. Their results were presented at the mid-term workshop ([KPI7/KPI9](#)).

- Stage 4 - **Evaluation, Dissemination, and Final Adjustments** - The final project stage included platform usability evaluation ([KPI10](#)), stakeholder engagement, and final dissemination.
 - **NOVARECKON** organized and hosted a mid-term workshop, coordinated expert panel participation, and synthesized feedback into the project's final adjustments. They also managed the SUS evaluation and communications through newsletter and other media.
 - **Mind in a Box** supported the evaluation by assisting with user testing and preparing technical summaries for dissemination. They also held the second workshop of the project at the World Summit AI 2025 in Montreal, helping position GALICIA within global AI safety and LLMOps discussions.

Open-Source Communities: *if your project involved interactions with open-source communities and there have been Memorandums of Understanding (MoUs) signed, please provide details on these agreements. Additionally, explain how your project has integrated with open-source communities, including any meetings or collaborations that have taken place during the experimentation phase.*

Before the end of the project, we have released all software developed within the project under an open-source license, specifically choosing **GPL**, to ensure both accessibility and usability by the community. This licensing not only allows others to use, modify and distribute our software freely, but also encourages collaborative development and innovation.

Galicia includes automated processes for code generation, formal model generation and test case generation based on user-submitted requirements. The system is designed to iteratively refine the generated code until all test cases pass, or a maximum number of iterations is reached.

The full codebase will be made publicly available on the **GitHub** platform <https://github.com/halservice/galicia-sargasso> and is developed in **PHP** using the **Laravel** framework. The project uses Laravel **Sail** and **Docker** to streamline the development environment and ensure consistency across different machines, with Docker containerizing the application and its dependencies for easier deployment. On the front end, **Livewire** enables the creation of reactive and dynamic interfaces directly within Laravel.

The platform is organized into Laravel modules such as models, traits, actions and several view modules such as configuration settings, system logs and statistical dashboards.

The backend is powered by a **PostgreSQL** database; the core structure includes:

- **Users and user settings:** these tables store fields like username, email, hashed password and a boolean `is_admin` field (default: false). Admin users are granted access to the "Export All" feature in the logs page. While the user setting saves the last settings for each user.
- **Generated codes:** stores user requirements, the corresponding generated code, system messages and the LLM used. It includes a nullable foreign key to the `formal_models` table, depending on whether the generation process started from code or model.
- **Formal models:** contain the user ID, initial requirement, generated formal model, model tool used and LLM reference. It includes a nullable foreign key to the `generated_codes` table, depending on whether the generation process started from

code or model.

- **Validated code:** includes the validation process output (as JSON), test cases and results, the final validated code, the maximum number of iterations attempted and polymorphic references to either code or model generation, enabling full traceability in both directions.

The final GitHub repository includes **setup instructions**, a **database structure** overview and the **GPL license** file to comply with open-source best practices. Although the project did not involve any direct collaboration with specific open-source communities during the experimentation phase, it is fully aligned with open-source principles and aims to encourage future contributions and integrations from the wider community after its release.

In line with our original commitment to actively engage with the open-source community, we have undertaken several outreach and dissemination efforts during the experimentation phase of the GALICIA project. Notably, we organized a [dedicated workshop in Italy](#) on 4th of March, 2025, which focused on presenting the platform's technical foundations and its open-source development approach to a community of developers, researchers, and practitioners. This event facilitated valuable exchanges and early feedback on the project's architecture and usability. Furthermore, GALICIA was presented during a [workshop held at the World Summit AI in Montreal](#), providing international visibility and fostering connections with a broader network of AI and software engineering professionals. Additionally, the infrastructure used in the GALICIA project, including its LLMOps and the proposed context-aware Retrieval-Augmentation Generation (RAG) workflow, was disseminated at ASPAI 2025 in Austria through a [scientific paper](#) and presentation, further contributing to the academic and technical discussion around the platform. These engagements reflect our intention to establish an active dialogue with the open-source ecosystem and lay the groundwork for future collaboration, contributions, and adoption. While no formal collaborations with open-source communities were established during the experimentation phase, these initial activities have helped position GALICIA as a transparent and community-oriented initiative. We anticipate expanding our involvement post-release through participation in forums, webinars, and further workshops, as initially outlined in the project proposal.

Impact analysis and/or feasibility study

The present section is where you should assess the results and potential of your project. Here, you should evaluate the outcomes, benefits, and challenges faced during the course of your experiment. It's the moment to showcase the tangible impacts and, if relevant, the feasibility of scaling your project for broader adoption. This section holds the key to demonstrating the value your project brings to the table and the expected results.



Kindly adhere to the outlined scheme while formulating your report adding any other relevant information:

Expected results. Comparison of what has been achieved over the program duration, and what was planned at the beginning.

At the outset, GALICIA aimed to explore the integration of generative AI with formal methods to improve the reliability and security of internet applications. The planned deliverables—such as the platform prototype, demonstrators, and user feedback—were completed within schedule, and KPIs like usability validation (KPI10) and workshop engagement (KPI9) confirmed strong stakeholder interest. Productivity gains through AI support aligned with expectations, while the extent of human oversight required was greater than anticipated, highlighting the importance of hybrid workflows. Overall, the project validated its core hypotheses and provided a solid foundation for further development and adoption.

*Describe the **progress of the project** so far towards delivering scientific impact, based on its objectives and towards delivering impact in any of the following fields (if applicable): scientific, economic, societal or industrial production or processes.*

GALICIA has made meaningful progress toward delivering scientific impact by demonstrating that generative AI can be effectively combined with formal methods to produce more secure and verifiable software systems. The project's three core use cases—in IoT, mathematical modeling, and network protocols—offered practical proof-of-concept that hybrid development workflows are feasible and beneficial. Scientifically, GALICIA contributes to emerging research on AI explainability, formal verification, and human-in-the-loop development strategies. On an industrial level, the demonstrators provided insights into how such techniques can be integrated into existing software engineering pipelines. Societally, the project responds to growing concerns over AI reliability by promoting transparent and traceable development methods. Early stakeholder feedback confirmed a strong demand for tools that balance innovation with trust and accountability.

*Report on changes to the **expected impacts** presented in your IMP (if any) and the effects on the project/need for adaptations. Where necessary, provide further details of your monitoring and evaluation strategy, including: references to baselines, benchmarks, assumptions used (with justification) as well as calculations performed to quantify the impacts.*

In line with the original Individual Mentoring Plan (IMP), the project employed a flexible and continuous monitoring throughout its lifecycle. Rather than relying solely on post-hoc evaluations, we implemented feedback loops at key stages—including milestone assessments, structured user interactions, and both mid-term and final workshops—to inform adaptive development. Benchmarks included qualitative indicators such as traceability across generated and validated components, iteration depth for code convergence, and usability feedback. While numerical baselines (e.g., pass rates for auto-generated test cases) were not

strictly formalized in the IMP, data from user interaction logs and informal evaluations helped identify patterns and priorities, which informed course corrections and longer-term planning. The GALICIA project concluded with a robust set of outcomes that align with the foundational objectives established in the original Individual Mentoring Plan (IMP). During the final phase, the team achieved a stable, open-source release of the full software stack, encompassing modules for AI-assisted code generation, formal model generation, iterative test case validation, and interactive user interface. The platform developed using PHP's Laravel framework, PostgreSQL, Docker, and Livewire, is openly available under a GPL license on GitHub, ensuring its accessibility for community reuse and further experimentation.

These final deliverables consolidated GALICIA's core innovation: a hybrid approach that integrates large language models (LLMs) with formal verification tools in a transparent and traceable software engineering workflow. While early aspirations aimed at fully automated code synthesis were tempered by practical limitations, this pivot resulted in a more resilient architecture that blends generative automation with structured oversight. The system iteratively refines outputs based on user-submitted requirements and validation results, and supports both code- and model-driven pathways. Importantly, all transformations—requirements to code, code to model, test validations—are recorded and linkable, ensuring traceability across the entire software lifecycle.

During the final months, emphasis was also placed on usability and user interaction. Frontend modules for system logs, statistics, and configuration were refined following user feedback, and workshops in Italy, Canada, and Austria provided vital feedback loops. These events not only confirmed the platform's technical viability but also validated the design decisions around explainability and traceable generation flows. While no formal community collaborations were initiated during the project runtime, these outreach activities laid a promising foundation for broader adoption.

Looking ahead to the next 12 months, the project's architecture and modular design offer strong potential for extension into additional domains where verifiable, AI-assisted development could yield significant impact. Candidate areas include regulated industries—such as medical devices, energy systems, telecommunications infrastructure, and mobility services—where formal compliance and system robustness are not optional but critical. Experimentation in these sectors will likely require adapting GALICIA's model to domain-specific standards and possibly integrating new verification backends or front-facing user guidance tools.

Additionally, the team has identified several methodological and technical areas for continued exploration. These include:

- **Semantic mapping between natural language requirements and formal specifications**, to reduce ambiguities in early-stage input.
- **Specializing code generation for specific domain constraints**, with Retrieval-Augmented Generation (RAG) being generally recognized as an appropriate solution to address such requirements.
- **Support for domain-specific modeling languages (DSMLs)**, enabling smoother adoption by professionals in vertical sectors (using the RAG from the previous bullet).



- **Enhanced explainability features**, such as visual trace maps and inline commentary generated alongside model transformations or code rewrites.
- **Collaborative features and role-based workflows**, especially relevant for regulated industries with multi-actor compliance chains.

It is expected that these enhancements will mature incrementally over the coming year, contingent on external contributions, partner alignment, and access to targeted pilot domains.

From a strategic standpoint, the long-term perspectives of the three core contributors diverge yet complement each other, which opens a pathway for differentiated exploitation of the GALICIA platform:

- **Novareckon**, with its strength in RTD project design and stakeholder engagement, is focused on expanding the platform's footprint within European innovation ecosystems. Its medium-term strategy involves integrating GALICIA into demonstrator scenarios targeting public-private partnerships, cybersecurity certification, and critical infrastructure resilience.
- **HAL**, the primary technical developer, sees potential for continuous refinement and reuse of the architecture in formal software development contexts, particularly in scenarios that demand robust verification and test automation. HAL aims to nurture the platform as a reference implementation for LLM/formal tool integration and to attract contributors through open-source governance and development challenges.
- **Mind in a Box**, by contrast, brings a strong applied AI background and is more oriented toward usability, market-fit, and interface innovation. Their likely focus in the year ahead will be on embedding GALICIA's capabilities into broader cognitive systems and adaptive interfaces, potentially offering customized solutions for SMEs in healthcare, logistics, and smart services.

In conclusion, GALICIA has successfully transitioned from a conceptually ambitious vision to a working prototype demonstrating hybrid AI-formal modeling in a transparent, traceable, and user-accessible manner. The achievements made in the final project phase lay a firm technical and strategic groundwork. With a committed open-source posture, active stakeholder engagement, and complementary long-term visions from the three core partners, the project is well-positioned to evolve into a scalable, cross-domain solution for trustworthy AI-assisted software engineering.

Intention to Create Startups/Spin-offs: *for projects originating from universities, explain if there is an intent to create startups or spin-offs related to the exploitation of the project and the steps taken in this direction.*

While GALICIA did not originate directly from a university setting, the collaboration between NOVARECKON and Mind in a Box fostered a fertile ground for future entrepreneurial developments. Discussions initiated during the final stages of the project suggest interest in exploring commercialization pathways, particularly for domain-specific applications of the

hybrid AI-formal methods platform (e.g., in secure IoT or regulatory-compliant software sectors). No formal steps toward the creation of a startup or spin-off have yet been taken, but both partners are evaluating potential business models, licensing strategies, and market opportunities. Further investment and stakeholder engagement would be necessary to assess the viability of a dedicated venture in the near future.

*Have there been any **new jobs** created directly or indirectly as a result of the programme? If so, please provide data on the number of new jobs created.*

Yes, the programme has indirectly led to the creation of one new job at HAL Service, the company that participated in the project by contributing a use case.

Is the team effectively maintaining the EU-US and EU-Canada partnerships by upholding existing agreements and ensuring the possibility of sustaining these connections in the future?

Yes, the EU-Canada team has remained active throughout the programme. Regular exchanges and collaborative activities have helped maintain strong communication and alignment with the Canadian partners. Moreover, the shared interest in continuing joint research initiatives and innovation provides a solid foundation for sustaining and potentially expanding these connections in the future. The relationships built during the project are expected to support further cooperation beyond the programme's formal end.

Publications

In the present section you should provide information on individual publications made during the participation in the programme. In addition to listing individual publications, you should also provide a brief overview of their open-source contributions. This should include a list of any open-source projects or developments that were made during the programme, as well as links to any online repositories where these projects can be found.

Public Deliverables

- KPI4 – Platform logical architecture and requirements (February 2025)
- KPI6 – Platform Delivery
- KPI7 – Final Project Results Report (March 2025)
- KPI9 – Workshop Proceedings (April 2025)
- KPI10 – Platform Evaluation Results (April 2025)
- KPI13 - Platform Verification and Evaluation

Communication Outputs

- KPI2 – First GALICIA Newsletter (December 2024)
- KPI8 – Second GALICIA Newsletter (April 2025)
- KPI12 - Terms of Reference of the final Workshop (April 2025)
- KPI11 - Third GALICIA Newsletter (June 2025)

Publications:

- Draft paper submitted to ANIPLA for their journal [Automazione e Strumentazione](#).
- [Mind in a Box Presentation at AI Conference in Montreal](#)
- [Scientific paper published at ASPAI 2025 in Austria by Mind in a Box](#)

All the above documents are published on the GALICIA web site (<https://www.galicia-project.eu/>)

