GALICIA

March 2025

Alberto Stefanini (ed., Novareckon) Alberta Bertin (Novareckon)

Franco Alberto Cardillo (CNR) Jérémie Farret (Mind in a Box)

Lorenzo Vandoni (HAL Service)

KPI9 – Proceedings of the mid-term GALICIA workshop

Project Name: GALICIA	
Report Title: KPI9 – Proceedings of the mid-term workshop	
Authors: Alberto Stefanini (ed.) and Alberta Bertin (Novareckon), Franco Alberto Cardillo (CNR), Jérémie Farret (Mind in a Box), Lorenzo Vandoni (HAL Service)	
Revised by: Alberto Servida (UniGE e ANIPLA)	
Date: 05/03/2025	
Version: 1	
Distribution: Public	

TABLE OF CONTENTS

1 INTRODUCTION

The GALICIA project aims to establish a platform for the automated verification of Large Language Model (LLM)-generated code against predefined test cases, with a strong focus on automation and industrial applications in the telecom sector. By encompassing the complete lifecycle—from natural language functional requirements and security specifications to formal verification—GALICIA seeks to bridge the gap between AI-generated code and compliance with user-defined requirements and industry standards. The project's broader ambition is to analyze the reliability of Generative AI in software development and contribute to building trust in AI-driven automation.

2 OBJECTIVES AND AGENDA OF THE WORKSHOP

The workshop was a key milestone in the development and refinement of the GALICIA platform, structured around the key outcomes of the stakeholder survey. It provides an opportunity to consolidate initial findings, present case studies, and outline future directions based on the challenges and priorities identified through the survey:

- 11:30 – 11:45 Alberto Stefanini: **Workshop objectives**
- 11:45 – 12:15 Franco Alberto Cardillo (CNR) **Formal methods in the era of Large Language Models**
- 12:15 – 12:45, Lorenzo Vandoni, **the platform**
- 12:45 - 13:15 L. Vandoni and A. Stefanini, **a set of study cases**
- 14:00 – 14:30 Jeremie Farret (Mind-in-a-Box), **Galicia's Hybrid LLM infrastructure: Differentiators and Commercial Applications,**
- 14:30 – 16:00 **Round Table Discussion**

In summary, the key objectives of the workshop were:
- Presentation and Discussion on the GALICIA Platform
- Presentation and discussion of several initial case studies
- The discussion in a Round Table of:
    - Specific technical objectives and opportunities for GALICIA development
    - An initial RoadMap to further develop GALICIA with contribution from the current stakeholde

The agenda has been designed to foster technical discussion and stakeholder engagement, ensuring that GALICIA evolves in a way that is both practical and directly aligned with the most pressing issues in cybersecurity, AI-driven automation, and compliance frameworks

3 THE SURVEY

The consultation, launched at the end of October 2024 will be closed for end of Survey Methodology.

The survey management and collected data are handled through Google Forms.

The link to the survey was sent to direct contact list of the project partners, and it was shared through social media, and Industries association that share it to their associates. The survey was

carried out to gather insights into the challenges and opportunities in verifying AI-generated code, as well as stakeholder perspectives on automated testing and compliance. Its outcomes played a crucial role in shaping the GALICIA platform requirements with respect to market expectation and establishing the ToR of the workshop, ensuring that its scope, objectives, and key discussion themes align with the most relevant issues identified. The ToR now serves as a strategic framework to guide stakeholder engagement, facilitating discussions on cybersecurity, AI-driven automation, and compliance verification in a way that directly addresses the survey's findings. The workshop outcomes will serve as a baseline to refine GALICIA's approach, incorporating feedback from industry and research communities to ensure that the project remains aligned with emerging challenges in digital resilience. With the support of NGI SARGASSO, the event will bring together key partners—including Novareckon, Mind in a Box Inc., and Hal Service—to demonstrate and evaluate the potential of automated verification in real-world applications. This report provides a foundation for these discussions, offering a comprehensive view of the project's current state and its future trajectory.

3.1 Survey Objectives

As part of the GALICIA project, a consultation was launched that involved users of

innovative SMEs and Start-ups, Research Centers, Large companies, Industries Associations, and Innovation Hubs, European and Canadian, to learn about their experience and their point of view regarding the innovation prospects of the project.

3.2 Survey Results

3.2.1 Geography of respondents and profile of the organisations

The total number of responders on the 8th of January 2025 are 64.

Most of them are from European countries, in order of numerousness Italy, flowed by Spain, France, Greece, and Belgium, while from Canada responders are 9% of the total.

Within Italy the distribution is among 11 regions over 20.

Respondents are representative of innovative SMEs and Start-ups, Research Centers, Large companies, Industries Associations, and Innovation Hubs. Most respondents are from Small Medium organisations (71.4%), Lage Enterprises are represented by 28.6%.

3.3 KEY INSIGTHS

The results of the GALICIA survey highlighted the significant interest and growing reliance on generative AI tools across various sectors, particularly within SMEs, large enterprises, and research centres. Respondents widely recognize the potential of generative AI in streamlining software development processes, with tools like ChatGPT, Google Gemini, and Microsoft Copilot being prominently used. However, this rapid adoption also underscores the pressing need for robust verification and security mechanisms to address the risks associated with AI-generated code.

Key insights from the survey include the following:

- **Adoption of Generative AI**: Most respondents (79.4%) are integrating generative AI into their organizational workflows, particularly for software development (60%), where Python leads as the preferred language.

- **Importance of Cybersecurity**: The survey revealed that security flaws in AI-generated code are a frequent concern, with 66% of respondents emphasizing the need for tools to ensure code quality and compliance.

- **Demand for Validation Tools**: 80% of respondents expressed the necessity for tools that can verify AI-generated code for correctness, security, and compliance, with anticipated benefits such as time savings (84%), cost reduction (58%), and improved security (56%).

These findings affirm the relevance and timeliness of GALICIA's mission to deliver an automated platform for validating AI-generated code. By addressing the identified gaps in trust, compliance, and efficiency, GALICIA is well-positioned to serve a critical role in advancing the secure and reliable adoption of generative AI technologies across industries. As the survey progresses and additional feedback is collected, these insights will further refine GALICIA's development priorities and strengthen its alignment with user needs and expectations.
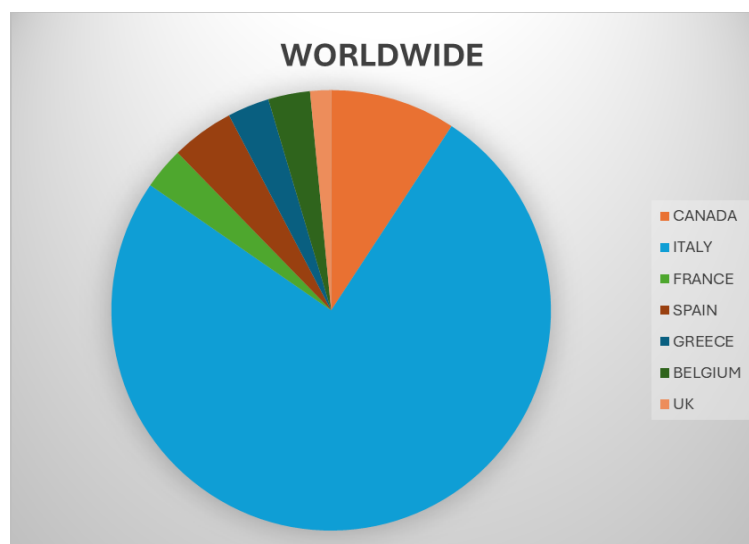

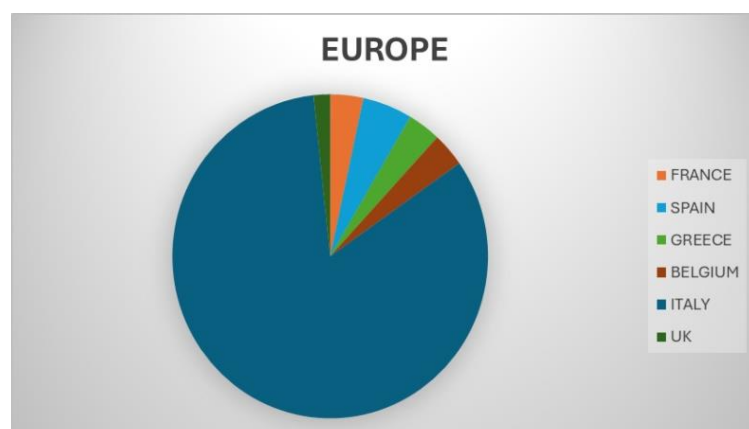*Figure 1 - Survey Data Worldwide, Source: GALICIA Survey data from Google Forms.*

Figure 2 - Survey data Europe, source: GALICIA Survey data from Google Forms.

4 FORMAL METHODS IN THE ERA OF NEURAL LANGUAGE MODELS

Franco Alberto Cardillo (CNR) presents an overview of formal methods in software engineering, discussing code verification and validation. He discusses the use of formal languages to express requirements and interaction with language models such as ChatGPT to generate specifications. Cardillo highlights the challenges in using these tools, such as scalability issues, and explores the potential of language models in assisting domain experts at various stages of software development.

4.1 Large Neural Networks

Franco Alberto Cardillo began his talk with an introduction to the role of formal methods in software verification and validation (V&V). While these methods provide proofs of software correctness, they require a high level of expertise, which is both rare and expensive. This limitation fuels the interest in AI-driven approaches to software V&V, which could have a significant impact on this crucial stage of the software life cycle, enabling more companies to develop safe and robust code. However, Cardillo emphasized that current off-the-shelf LLMs still struggle to generate robust, safe and correct code, often producing plausible but flawed results.

The presentation then shifted to describing how LLMs are trained and specialized to Code LLM. Cardillo showed how a fine-tuning procedure, tailored for a specific application domain or task, can improve the ability of these models to detect errors or generate formal specifications. He presented recent research where LLMs were integrated in hybrid systems capable of correcting bugs or verifying some formal properties of algorithms. The presentation concluded with a cautionary message on how to correctly interpret the output of models, especially those trained on unknown or opaque datasets, raising some concerns about their ability to generalize to domains far from those characterizing their training set.

5 PRESENTATION AND DISCUSSION ON THE GALICIA PLATFORM

The session opens with a comprehensive presentation of the GALICIA platform, highlighting its key capabilities, architectural foundations, and areas of ongoing refinement. This overview set the stage for discussions on how GALICIA can support industries in achieving automated compliance, cybersecurity risk assessment, and software verification through AI-driven approaches. Special attention will be given to recent advancements in integrating formal methods with machine learning techniques, a topic of relevance given the growing demand for robust verification frameworks in mission-critical environments.

In this session, a demo of the currrent Alpha version of the Galicia platform has been shown, using two small examples in C language.

GALICIA

CASE STUDIES: INITIAL INSIGHTS AND LLM-DRIVEN ANALYSIS

Following the platform presentation, the discussion will transition into an analysis of several initial case studies that have been conducted using GALICIA's capabilities. These case studies will illustrate practical applications in different domains, ranging from telecommunications and industrial automation to cybersecurity risk evaluation. Each study will highlight specific technical challenges addressed by GALICIA, as well as preliminary outcomes that demonstrate its effectiveness in improving compliance efficiency and security assurance.

Additionally, a separate segment will be dedicated to case studies that leverage a single specific Large Language Model (LLM) to automate certain processes within GALICIA. These cases provided insights into the feasibility and potential impact of LLM-driven functionalities, such as code analysis, vulnerability detection, and compliance cross-checking. The findings from this exploration informed the further discussions on the broader role of AI in automating traditionally complex verification and validation tasks.

The GALICIA's methodology involves a structured workflow that bridges NL specifications with formal verification techniques:

1. **Formal Model Creation:** Derive formal specifications from the user-provided natural language functional requirements (NLFR)
2. **Code Generation:** Use LLMs to generate source code from NLFR or generated code using several formal methods
3. **Compliance Verification:** Employ model-checking techniques to ensure the formal model adheres to user-specified functional and security requirements
4. **Iterative Refinement:** Revise the generated code based on verification feedback
5. **Code Compilation and Validation:** Translate formal specifications into executable code, maintaining consistency across languages
6. **Deployment Testing:** Validate the refined code in a real-world environment to confirm robustness and adherence to requirements

5.1 Case studies landscape

Basically, it encompasses:

- fictitious cases from known mathematical and geometrical problems such as the Fibonacci sequence and Euclid's geometry
- Specific functions on HAL's WiC advisor
- The implementation of a portion of the TCP/IP protocol, the handshake mechanism was implemented

5.2 Proving a Euclid Theorem

Euclid did stated his axioms in advance, in form of:

- 23 **Definitions** (e.g,, 1 A point is that which has no part., 2 A line is breadthless length)
- 5 **Postulates** (e.g., 1. (it is possible) to draw a straight line from any point to any point)

- 5 **Common notions** (e.g., 5. The whole is greater than the part)

From the above you may derive the formal Model of Euclid's geometry in any formalism of choice, e.g., a set of clauses in formal logic:

- "it is possible to draw a straight line from any point to any point" could be written in first-order logic as: $\forall A, B \exists L(Line(L) \land PassesThrough(L,A) \land PassesThrough(L,B))$

From the above it is possible to mathematically build an Equilateral Triangle:

- Points are represented in terms of a couple of coordinates and relations are defined between them, like distance and equality

A Prolog code excerpt is given below:

- % Define points as facts (assuming A and B are known)
- point(a, 0, 0).
- point(b, 4, 0).
- % Define a predicate to calculate distance between two points
- distance((X1, Y1), (X2, Y2), D) :-
-     D is sqrt((X2 - X1)^2 + (Y2 - Y1)^2).

5.3 A set of test cases on the WiC Advisor

The WiC Advisor is a proprietary interface of HAL Service, to monitor, manage and maintain a telecom infrastructure. The test cases encompassed

- The **CRUD** functions (Create, Read, Update, Delete) enable a customer data base resource management, e.g. querying, resource creation, efficient and secure handling of IoT services
- **Check coverage**: [1]enter an address, e.g. Piazza Garibaldi 6 Milan, verify that it is correct and complete, connect to web services provided by local operators (e.g. Tim. Open Fiber, FastWeb), and show the services available at that address, together with the maximum speed (e.g. 100Mbps) and the price (e.g. €19.90/month).
- **Quotes:** a wizard guiding the user to create a quote, allowing him to enter customer name, a set of required products and fheir price
- **Ticket:** enter a request for technical assistance, call technical support, record the taking of charge, and the positive or negative outcome

5.4 The TCP/IP handshake mecanism

The case study focused on the TCP three-way handshake protocol, basically on:

- an Event B model generated by GPT
- Its verification in a Python implementation

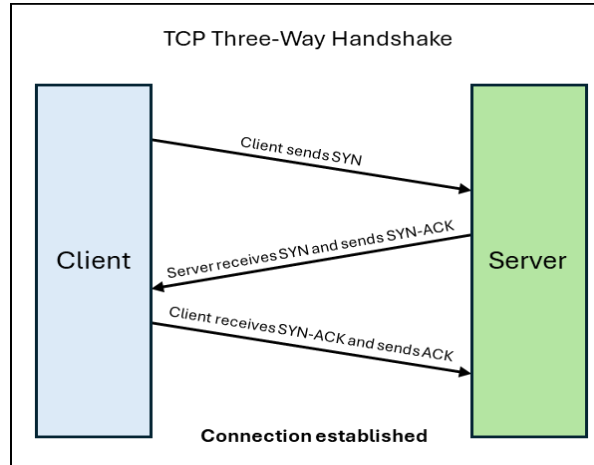[1] The red label denotes work in progress.

Figure 3 - The TCP/IP Handshake scheme.

This sequence ensures reliable communication between sender and receiver, it typically encompasses the following steps:

1. the Sender sends a message to the Receiver
2. Upon receiving the message, the Receiver validates its integrity (e.g., checksum)
3. If the message is valid, the Receiver sends an Acknowledgment (ACK) to the Sender
4. The Sender:
   - Waits for the ACK within a predefined timeout
   - If the ACK is received, considers the message successfully transmitted
   - If no ACK is received (timeout), retransmits the message
5. This process repeats until the message is successfully acknowledged or a maximum retry limit is reached

The Formal model was generated by GPT in a co-creative process, the LLM provides indeed extensive user support. Among other, ChatGPT suggested to use the Event B formalism based on the problem features:
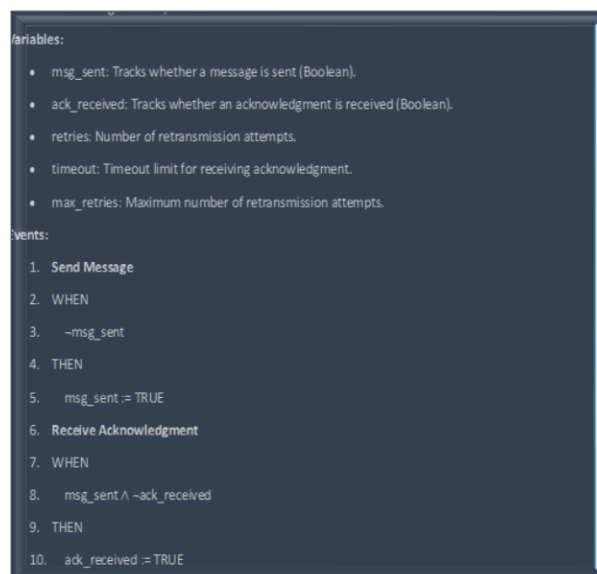

Figure 4 - Excerpt of the Event B model of the handshake protocol.

This model was then implemented in Python, a programming language extensively used by developers, as testified by our stakeholder survey. The resulting code is very similar to Event B, note how events are marked by printouts:

```
# Property 1: Message must be sent before acknowledgment
assert msg_sent, "Acknowledgment cannot be received without sending a
message."

print("Waiting for acknowledgment...")
time.sleep(timeout)

if receive_ack():
    ack_received = True
    print("Acknowledgment received. Process complete.")
    # Property 2: Acknowledgment ensures termination
    assert retries <= max_retries, "Retries exceeded after acknowledgment."
else:
    retries += 1
    print(f"No acknowledgment. Retrying ({retries}/{max_retries})...")
    msg_sent = False
```

Figure 5 - Python implementation of the Formal Model.

For a number of commercial domains and markets, Python would need to be substituted with other scripting languages, and specific know-how would become necessary to produce and validate code. Jérémie remarked that this consideration implies the necessity to be able to enrich the accessible knowledge for GALICIA. The Mind in a Box infrastructure for LLMOps supports conventional enrichment strategies such as retrieval-augmented generation (RAG).

6   MIND-IN-A-BOX CONTRIBUTION

Jérémie Farret presented on behalf of Mind in a Box, discussing the infrastructure supporting the project and its commercial applications, highlighting potential value for customers in areas like cybersecurity and IoT. He also shared his experience with commercialization programs in Canada, emphasizing the importance of supporting multiple languages and the challenges of cloud infrastructure for safety-critical applications. Finally, Jérémie suggested targeting niche markets that can afford premium solutions to avoid competing directly with large tech companies offering cheap or free AI services.

6.1 Development

Jérémie to explore international collaboration programs with the International Chamber of Commerce in Quebec and share promising avenues with all partners.

Valentin to provide more details on emergency vehicle fleet management use cases if needed.

Galicia team investigate the potential of retrieval augmented generation for addressing language support and specific customer needs.

Galicia team to consider focusing on niche markets or verticals that have safety and regulatory concerns and are willing to pay a premium for on-premises solutions.

Alberta to further explore the possibility of targeting small and medium enterprises for Galicia's business model.

Galicia team to evaluate the strategy of offering a public platform for visibility while targeting specific verticals for revenue generation.

6.2 Galicia's Value in Script Automation

Jérémie presents on behalf of Mind in a Box, one of the project partners, explaining the infrastructure supporting the project and how it differs from typical GPT infrastructure. He discusses commercial applications and use cases where Galicia could provide value to their customers, particularly in areas like script generation for data ingestion, cybersecurity, IoT, and AI operations. Jérémie highlights how Galicia could help automate and simplify scripting tasks across their various solutions and customer deployments, potentially reducing costs and improving efficiency.

6.3 Python Scripting's Role in AI & Data

Jérémie discussed the heavy use of scripting in various fields including artificial intelligence, data operations, and analytics. He highlighted Python as a universally used scripting language in these areas. Jérémie also emphasized the importance of a system like Galicia, which produces verified and reliable scripts from prompts, particularly in high-stakes applications involving public safety and security.

6.4 Commercialization Programs and International Collaboration

Jérémie shares his experience with commercialization programs in Canada, highlighting three types of grants: follow-up grants from the same program that funded initial research, regional programs, and domain-specific programs targeting particular industries. He mentions the Scale AI program as an example but notes it wouldn't be directly applicable to their current situation. Jérémie also discusses ongoing efforts to explore international collaboration programs between Canada and the European community, promising to share any promising avenues with all partners once identified.

6.5 Language Support in Bilingual Markets

Jérémie discusses the importance of supporting multiple languages, especially in bilingual countries like Canada, and highlights the advantages of the Money in the Box infrastructure for language support. He emphasizes the significance of retrieval-augmented generation for addressing specific customer needs and complementing language support. Jérémie also points out the challenges of cloud infrastructure for safety-critical applications due to regulation and cost concerns. The discussion then shifts to market strategies, with Jérémie suggesting targeting niche markets that can afford premium solutions to avoid competing directly with large tech companies offering cheap or free AI services.

7 ROUND TABLE DISCUSSION: OBJECTIVES, OPPORTUNITIES, AND ROADMAP

The event was concluded by a round table discussion, where stakeholders engaged in a structured dialogue about the strategic and technical objectives for GALICIA's further development. Key points for discussion were:

- The identification of specific technical objectives that can drive the next phase of GALICIA's evolution, including areas for refinement and expansion.

- Opportunities for industry adoption and collaboration, particularly in sectors where compliance automation and AI-driven validation can yield significant benefits.

- The initial development of a roadmap that will guide GALICIA's future trajectory. This roadmap will incorporate insights from current stakeholders, ensuring that the platform continues to align with real-world needs and emerging regulatory trends.

By synthesizing these discussions into a clear set of action points, the event contributed to shaping GALICIA into a more robust and adaptable solution. The engagement of industry experts, research institutions, and SMEs was instrumental in ensuring that GALICIA remains at the forefront of AI-driven compliance and cybersecurity innovation.

Figure 6 - Technical issues pertaining the use of LLMs.

Large Language Models (LLMs) present both opportunities and challenges for the Galicia project, particularly in their interaction with formal methods. It is important carefully evaluating the performance of AI models, highlighting the phenomenon of overfitting and the need for accurate verification. We conclude by emphasizing the potential of AI in assisting formal software verification, but a cautious approach is needed, and one must be aware of the current limitations of large language models. Formal models remain conceptually tied to computer science foundations, and it may be difficult to understand and take profit from formalisms, both for the end-user and the LLM which may struggle and with the formalism rigor due to their probabilistic nature. to ensure reliability in critical applications.
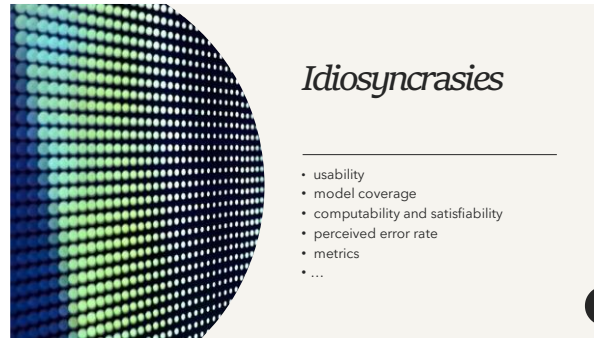
Figure 7 – Main issues discussed on difficulties of using LLMs.

A key difficulty relates to the fact that for most end-users, formal models are still rather arcane:

- Conceptually, they are strongly related to computer foundations & theoretics
- Moreover, depending on the way the model is expressed, it may be more natural to resort to a declarative approach (such as logics) or a procedural one (like finite state automata, Petri nets etc.)

But the key point, on that respect, is that, once the end-user interiorizes and understands the formalisms are understood, it is difficult to judge *a-priori* whether they are applicable to a large problem. Scalability remains a key concern: can LLMs effectively handle large-scale formal reasoning, or do they break down when applied to complex systems? Addressing these issues requires balancing automation with human oversight.

Figure 8 - How to promote RTDI.

Potential developments may concern:

AI Limitations and Caution

Novareckon presents an overview of the challenges and limitations of artificial intelligence, focusing on the problems of generalizing and interpreting machine learning models. It discusses the importance of carefully evaluating the performance of AI models, highlighting the phenomenon of overfitting and the need for accurate verification. Novareckon concludes by emphasizing the potential of AI in assisting formal software verification but points out that a cautious approach is needed and one must be aware of the current limitations of large language models.

New technologies and legal issues

The meeting covers a variety of topics, including the cost of new technologies, optimizing employment demands, and legal issues. Economic issues are also discussed, with references to the

Bank of Italy and Kazakhstan. Novareckon mentions the importance of managing public opinion and paying attention to new ideas. The meeting concludes with discussions on international collaborations and economic preparations.

Figure 9 - Dissemination and Exploitation channels.

Further case studies

To promote GALICIA through further case studies, the project should focus on demonstrating its applicability across diverse industrial domains, such as energy, manufacturing, and telecommunications. Engaging with SMEs and large enterprises to validate the platform in real-world scenarios can provide valuable feedback and strengthen its credibility. Publishing case studies in academic conferences, industry forums, and standardization bodies will enhance visibility and encourage adoption.

Making the platform available to the public requires a well-structured deployment strategy, including cloud-based or on-premise options. It is essential to establish clear access mechanisms, such as open-source licensing, subscription models, or partnership agreements, to attract users from both research and industry.

The Platform

Defining access tariffs should balance affordability with sustainability. A tiered pricing model could be implemented, offering free access for research and educational purposes while introducing premium plans for commercial users needing additional features, support, or integration capabilities.

Providing supporting materials is crucial for adoption. Comprehensive documentation, including user manuals, API references, and training modules, should be made available online. Interactive tutorials and workshops can facilitate onboarding, while FAQs and technical support channels can ensure continuous assistance for users.

Promotion of further development

To promote further developments, the GALICIA project should foster collaborations with research institutions, standardization bodies, and industry stakeholders. Establishing an open innovation ecosystem, encouraging third-party contributions, and integrating with emerging cybersecurity frameworks can ensure long-term growth. Additionally, seeking funding opportunities and aligning with European digital initiatives will help sustain and expand the platform's impact.

8 CONCLUSIONS

The GALICIA workshop, shaped by the insights gathered through the stakeholder survey, represents a pivotal step in refining the project's objectives and methodology. By structuring discussions around the key findings of the survey, the workshop ensured that GALICIA remains aligned with industry needs, regulatory developments, and the technical challenges of verifying AI-generated code. The survey not only helped define the scope of the event but also guided the formulation of its Terms of Reference, ensuring that the workshop addresses the most pressing issues in cybersecurity, compliance automation, and AI-driven validation.

The event's agenda, culminating in a structured roundtable discussion, was designed to facilitate meaningful engagement among industry experts, research institutions, and SMEs. The identification of technical objectives, industry adoption strategies, and the initial drafting of a roadmap will provide a concrete foundation for GALICIA's next development phase. The roadmap, enriched by stakeholder contributions, will serve as a guiding instrument to enhance the platform's capabilities while ensuring its adaptability to evolving regulatory and technological landscapes.

Moreover, the workshop not only consolidated existing research but also acted as a catalyst for future collaborations. By bringing together diverse stakeholders, the event strengthened GALICIA's position as a key enabler of trust and reliability in AI-driven software verification. Through continued engagement, the project will refine its methodologies, expand its case studies, and integrate best practices from industry and academia to maximize its impact.

Ultimately, this process will help GALICIA evolve into a robust, industry-ready solution, offering practical and scalable approaches to compliance verification in AI-generated software. The structured dialogue established in this workshop will be instrumental in ensuring that the project remains at the forefront of AI-driven cybersecurity and compliance innovation, fostering a more resilient and trustworthy digital ecosystem.

PROJECT CONSORTIUM: