Annales Mathematicae et Informaticae
63 (2026) pp. 65–72
DOI: 10.33039/ami.2026.06.006
URL: https://ami.uni-eszterhazy.hu

# Baseline review of Formal Methods and Foundations of Artificial Intelligence

**Gábor Kusper**

Eszterházy Károly Catholic University
kusper.gabor@uni-eszterhazy.hu

**Abstract.** Artificial Intelligence (AI) is advancing rapidly, yet many successful models remain opaque and provide limited assurance about reliability, safety, and failure modes. This motivates renewed interest in formal methods and foundational perspectives that can support trustworthy AI beyond empirical testing. This paper presents a baseline review of the selected papers volume of the inaugural International Conference on Formal Methods and Foundations of Artificial Intelligence (FMF-AI 2025), published as *Annales Mathematicae et Informaticae*, Vol. 61 (2025). The goal is twofold: (i) to map and summarize the first FMF-AI "snapshot" as a starting point for the Hungarian research ecosystem, and (ii) to define a reproducible baseline that can serve as a reference for measuring topical and methodological shifts in subsequent FMF-AI editions. The review clusters the twenty selected papers into five thematic groups and records their relative prevalence. In addition, it introduces simple baseline metrics that can be recomputed in future FMF-AI editions to observe structural changes in the research landscape. The main pattern is a clear imbalance between verification-oriented contributions and papers that primarily use AI methods in application or optimization contexts.

*Keywords:* FMF-AI, baseline review, Formal Methods and Foundations of Artificial Intelligence

2020 *Mathematics Subject Classification:* Primary 00A17, 68-02; Secondary 68T01

## 1. Introduction

Artificial Intelligence (AI) has achieved remarkable empirical success across a wide range of domains, yet this success is often accompanied by limited transparency,

Submitted: March 31, 2026
Accepted: June 30, 2026
Published online: July 17, 2026

weak interpretability, and uncertain failure modes. As AI systems become more pervasive, questions of reliability, accountability, and explanatory adequacy become increasingly central. In this setting, formal methods and foundational inquiry offer a complementary perspective: rather than relying exclusively on benchmark performance and empirical testing, they seek analyzable models, explicit assumptions, and, in some cases, guarantees about system behavior.

This perspective is particularly relevant for the International Conference on Formal Methods and Foundations of Artificial Intelligence (FMF-AI). The vision articulated in the conference foreword presents FMF-AI as a meeting point between theoreticians and practitioners, between formal reasoning and real-world AI applications, and between technical performance and the broader demand for responsible and explainable AI. It also frames the conference series as a growing community effort, rooted in collaboration, curiosity, and shared knowledge, and intended to develop into a recurring international forum.

FMF-AI 2025 produced two complementary publication outputs. In addition to the selected-papers special issue published in *Annales Mathematicae et Informaticae*, Vol. 61 (2025), the conference also resulted in a separate proceedings volume available on the conference website[1]. This dual publication context matters for the present study. The proceedings reflect the broader interdisciplinary profile of the event, spanning topics from mathematical foundations and large language models to educational and application-oriented themes, whereas the selected-papers volume offers a more compact and curated snapshot of contributions that are especially suitable for baseline comparison.

The purpose of this paper is not to rank the included papers or to evaluate them against a single technical standard. Instead, it provides a descriptive baseline review of the twenty selected papers published in the special issue. The central question is simple: when the first FMF-AI selected volume is viewed as a whole, what structural picture emerges from it, and what does that picture reveal about the balance between formal-methods-oriented work and the broader use of AI across domains?

More specifically, the review has two objectives. First, it maps and summarizes the thematic landscape of the FMF-AI 2025 selected papers. Second, it defines a small set of reproducible baseline descriptors that can later be recomputed for subsequent FMF-AI editions. The underlying assumption is that even simple counts and stable category assignments become informative when they are applied consistently over time.

The contribution of the paper is therefore threefold. It proposes a stable five-group thematic organization of the selected volume; it derives simple baseline metrics from this organization; and it offers an initial interpretation of the relationship between verification-oriented contributions and application-oriented AI research. The paper remains intentionally descriptive: its primary goal is to establish a usable reference point for later comparative work, particularly within the Hungarian research ecosystem to which several contributions in the volume are directly con-

[1] https://uni-eszterhazy.hu/fmf2025/proceedings

nected.

The remainder of the paper is organized as follows. Section 2 explains the scope and review design. Section 3 presents the five thematic groups. Section 4 summarizes the baseline metrics and their interpretation. Section 5 discusses limitations and transparency issues, and Section 6 concludes the paper.

## 2. Scope and review design

This review covers the twenty selected papers published in the FMF-AI 2025 volume of *Annales Mathematicae et Informaticae*. The paper adopts a descriptive, taxonomy-driven scoping-review mindset. Its goal is not exhaustive literature synthesis beyond the volume, but the creation of a reproducible baseline for this particular corpus.

For each paper, the review considers four descriptive dimensions: the addressed problem, the dominant method family, the main form of evidence, and the paper's explicit connection to trustworthy AI. The method family may be learning-based, optimization-oriented, symbolic or formal, or hybrid. The evidence may rely mainly on empirical evaluation, survey or user-study evidence, theoretical analysis, or formal guarantees. The trustworthy-AI connection may concern robustness, assurance, verification, safety-related reasoning, or a weaker indirect relation.

In the present baseline version, the primary emphasis is on thematic clustering and on a few structural metrics that can be computed in a straightforward and reproducible way. Since some papers could plausibly fit more than one perspective, each contribution is assigned to one primary thematic group. This choice sacrifices some nuance, but it enables stable paper counts and clear comparisons across later FMF-AI editions.

## 3. Thematic structure of the FMF-AI 2025 Volume

The selected papers can be organized into five thematic groups. These groups do not claim to be the only possible taxonomy; rather, they provide a stable descriptive structure that is broad enough to cover the whole volume while still highlighting meaningful differences among the contributions.

The number of groups was fixed at five for pragmatic methodological reasons. With only twenty papers, a finer taxonomy would easily produce several one-paper or two-paper categories, making the baseline unstable and difficult to compare across future editions. Conversely, a smaller taxonomy with only two or three groups would merge substantially different types of AI-related work, for example language technology, educational applications, and optimization-oriented contributions. The five-group structure therefore represents a compromise between interpretability, coverage, and future reusability.

### 3.1. Group I: Formal Verification of AI

Group I contains a single paper that explicitly combines empirical analysis with *formal verification* of robustness properties in neural network ensembles [14]. Within the selected volume, this contribution is the clearest realization of the conference's promise to connect formal methods with AI reliability. Its singular position is analytically important, because it marks the narrowest and most explicit core of what may be called verification-oriented FMF-AI research.

### 3.2. Group II: Testing and Reliability of AI

Group II gathers papers that address reliability concerns through empirical and engineering-oriented approaches rather than full formal verification. The covered topics include motion-enhanced video anomaly detection, AI-based interpretation of static human arm signals, comparative robustness and noise-sensitivity analysis in deep neural networks, and fault diagnosis based on textual system logs [2, 4, 11, 13]. These papers are closely related to dependability and trustworthy AI, but their contribution is primarily testing-, diagnosis-, or robustness-oriented.

### 3.3. Group III: Using AI for Text Analytics

Group III covers language- and text-related applications of AI. The papers in this group address toxic comment detection in Hungarian, syntactic comparison between human-written and AI-generated scientific texts, and the enhancement of a Hungarian conversational model [9, 19, 20]. Taken together, these contributions reflect both local-language priorities and broader questions raised by generative AI and language technologies.

### 3.4. Group IV: Using AI for Education

Group IV captures the strong educational and human-centered dimension of the volume. It includes automated fair team formation for STEAM activities using SMT, adaptive testing for programming proficiency, AI-based robotic applications in higher education, the impact of LLM use on algorithmic thinking, and a survey of AI-usage attitudes among Hungarian informatics students [1, 3, 6, 12, 15]. This group is especially important because it shows that FMF-AI 2025 is not restricted to technical verification questions; it also engages with pedagogy, human factors, and societal uptake.

### 3.5. Group V: Using AI for Optimization

Group V is the largest thematic cluster. It brings together papers in which AI, optimization, graph algorithms, or broader computational modeling play the central role. The topics include graph neural networks for refactoring P4 programs, the transition from rule-based to learned behaviors, multi-objective evolutionary scheduling, team coordination on graphs under risk, probabilistic models for sports

betting strategies, an improved Industry 4.0 reference architecture model, and community detection through overlapping label propagation [5, 7, 8, 10, 16–18]. Although these papers differ substantially in application area and methodological emphasis, they share the common characteristic that AI or computational intelligence is used as a constructive tool for solving domain problems.

# 4. Baseline metrics and discussion

The five-group taxonomy can be summarized with a simple paper-count distribution shown in Table 1. Even this compact representation already reveals an interpretable structural pattern.

**Table 1.** Baseline distribution of the FMF-AI 2025 selected papers.

| Thematic group | Number of papers |
|---|---|
| Group I: Formal Verification of AI | 1 |
| Group II: Testing and Reliability of AI | 4 |
| Group III: Using AI for Text Analytics | 3 |
| Group IV: Using AI for Education | 5 |
| Group V: Using AI for Optimization | 7 |
| Total | 20 |

A first aggregate metric emerges when Groups I and II are merged into a broader *verification-and-testing* side of the volume, while Groups III–V are merged into a broader *using-AI* side. Under this aggregation, the distribution becomes 5 versus 15, that is, a ratio of 1 : 3.

This ratio should be interpreted carefully. On the one hand, it indicates that the selected FMF-AI 2025 volume is dominated by papers that use AI methods in application, education, language, or optimization settings. On the other hand, it also shows that work directly addressing reliability is present, but concentrated mostly in empirical testing and robustness analysis rather than in explicit formal verification.

The most striking baseline observation is therefore that, despite the conference title, *explicit* formal verification appears in only one selected paper. This does not imply that the remaining papers are disconnected from trustworthy AI. Rather, it suggests that the strongest formal-methods interpretation of the conference theme is currently represented by a relatively small subset of the volume, while a much broader set of papers explores how AI can be applied, adapted, or evaluated in diverse contexts.

This imbalance should not necessarily be interpreted as a lack of relevance of formal methods. A more plausible interpretation is that rigorous verification techniques remain difficult to apply at scale in modern AI settings, whereas empirical testing, robustness studies, and application-driven AI methods are easier to deploy

across a broader variety of problems. In this sense, the present paper count distribution provides a useful baseline against which future FMF-AI volumes can be compared.

# 5. Limitations and transparency

The present review is intentionally narrow in scope. First, it examines only one selected-papers volume, so its observations should be read as a baseline description rather than as a mature characterization of the whole FMF-AI research direction. Second, the thematic grouping uses a single primary assignment for each paper, even though some contributions could reasonably belong to more than one category. Third, the current version focuses mainly on thematic distribution and a small number of structural metrics; later versions may add fuller coding of method families, evidence types, and trustworthy-AI dimensions.

A transparency issue should also be noted. The author is a co-author of two papers in the reviewed volume [6, 15]. For this reason, the present review deliberately avoids ranking, endorsement, or evaluative scoring of individual contributions. Its aim is descriptive organization rather than comparative judgment.

# 6. Conclusion and future work

This paper transforms the FMF-AI 2025 selected volume from a collection of individual papers into a structured baseline that can support later comparison. The main outcome is a stable five-group thematic map together with a small set of reproducible metrics, most notably the paper-count distribution across groups and the $1:3$ ratio between the verification-and-testing side and the broader using-AI side.

Beyond the descriptive contribution, the paper also highlights a substantive tension already visible in the first FMF-AI snapshot: the conference theme strongly invokes formal methods and foundations, yet the selected volume is much more strongly populated by papers that use AI in applied, educational, linguistic, and optimization-oriented contexts than by papers that perform explicit formal verification. Whether this pattern will persist or change in later editions is an open and interesting question.

Future work will extend the present baseline in at least two directions. First, the coding scheme can be refined by systematically analyzing method families, evidence types, and stronger or weaker forms of trustworthy-AI linkage. Second, the same taxonomy and baseline metrics can be recomputed for subsequent FMF-AI volumes, allowing the series to be studied longitudinally rather than only descriptively at a single point in time.

# References


[1] A. A. Adil, G. Kovásznai, M. Khalid: *Automated fair team formation in STEAM activities using Satisfiability Modulo Theories (SMT)*, Annales Mathematicae et Informaticae 61 (2025), pp. 1–14, doi: 10.33039/ami.2025.10.001.

[2] M. I. Almurumudhe, O. Hornyák: *Motion enhanced video anomaly detection using masked autoencoder and hybrid loss functions*, Annales Mathematicae et Informaticae 61 (2025), pp. 15–30, doi: 10.33039/ami.2025.10.015.

[3] A. Apró, T. Tajti: *An adaptive testing system for programming proficiency using Item Response Theory*, Annales Mathematicae et Informaticae 61 (2025), pp. 31–42, doi: 10.33039/ami.2025.10.018.

[4] M. Z. Bagladi: *Artificial Intelligence for interpreting static human arm signals*, Annales Mathematicae et Informaticae 61 (2025), pp. 43–54, doi: 10.33039/ami.2025.10.005.

[5] B. S. Csüllög, M. Tejfel: *Using GNN for refactoring P4 programs*, Annales Mathematicae et Informaticae 61 (2025), pp. 55–67, doi: 10.33039/ami.2025.10.021.

[6] I. Drămnesc, E. Ábrahám, L. Antal, P. Csóka, N. Fachantidis, T. Jebelean, G. Kusper, K. Papadopoulos: *Artificial Intelligence based robotic applications for higher education*, Annales Mathematicae et Informaticae 61 (2025), pp. 68–79, doi: 10.33039/ami.2025.10.009.

[7] R. Erdész, E. Troll: *Making the boss smarter: A journey from rule-based to learned behaviors*, Annales Mathematicae et Informaticae 61 (2025), pp. 80–93, doi: 10.33039/ami.2025.10.017.

[8] L. Á. Fazekas, K. Nehéz: *Multi-objective genetic and memetic algorithms in flexible flow-shop scheduling*, Annales Mathematicae et Informaticae 61 (2025), pp. 94–107, doi: 10.33039/ami.2025.10.011.

[9] P. Hatvani, Z. G. Yang: *Automated detection of toxic comments in Hungarian*, Annales Mathematicae et Informaticae 61 (2025), pp. 108–117, doi: 10.33039/ami.2025.10.007.

[10] A. Izsó, I. Harmati: *Solving the Team Coordination on Graphs With Risky Edges problem for nonzero self-loop weights*, Annales Mathematicae et Informaticae 61 (2025), pp. 118–128, doi: 10.33039/ami.2025.10.008.

[11] M. A. Khudhair, A. Fazekas: *A comparative study on the noise sensitivity of binary classification based on robust deep neural networks*, Annales Mathematicae et Informaticae 61 (2025), pp. 129–140, doi: 10.33039/ami.2025.10.006.

[12] S. Király, E. Troll: *Algorithmic thinking at risk? Exploring LLM use in computer science education*, Annales Mathematicae et Informaticae 61 (2025), pp. 141–155, doi: 10.33039/ami.2025.10.020.

[13] Á. Kiss, K. Nehéz, O. Hornyák: *AI-driven fault diagnosis from textual system logs*, Annales Mathematicae et Informaticae 61 (2025), pp. 156–170, doi: 10.33039/ami.2025.10.003.

[14] Á. Kovács, R. Gunics, G. Kovásznai, T. Tajti: *Soft voting robustness in neural network ensembles with empirical analysis and formal verification*, Annales Mathematicae et Informaticae 61 (2025), pp. 171–185, doi: 10.33039/ami.2025.10.002.

[15] G. Kusper, G. I. Mátyás, T. Balla: *We are not afraid of the wolf! – AI usage attitudes among Hungarian informatics students*, Annales Mathematicae et Informaticae 61 (2025), pp. 186–201, doi: 10.33039/ami.2025.10.004.

[16] J. G. Pál, C. Bíró: *Evaluating profitability in sports betting using probabilistic models and betting strategies*, Annales Mathematicae et Informaticae 61 (2025), pp. 202–214, doi: 10.33039/ami.2025.10.016.

[17] I. Sima, D.-M. Cristea, E.-M. Ciortea, L. B. Iantovics: *cRAMI 4.0 an Improved Reference Architectural Model for Industrie 4.0 (RAMI 4.0) based on a three-dimensional cubic lattice model*, Annales Mathematicae et Informaticae 61 (2025), pp. 215–228, doi: 10.33039/ami.2025.10.023.

[18] S. P. Tahalea, M. Krész: *Betweenness-driven overlapping label propagation community detection*, Annales Mathematicae et Informaticae 61 (2025), pp. 229–247, doi: 10.33039/ami.2025.10.012.

[19] E. B. Varga, A. Baksa: *Syntactic comparison of human and AI-written scientific texts*, Annales Mathematicae et Informaticae 61 (2025), pp. 248–260, doi: 10.33039/ami.2025.10.013.

[20] Z. G. Yang, Á. Bánfi, R. Dodé, G. Ferenczi, F. Földesi, P. Hatvani, E. Héja, M. Lengyel, G. Madarász, M. Osváth, B. Sárossy, K. Varga, T. Váradi, G. Prószéky, N. Ligeti-Nagy: *ChatPULI: Enhancement to the first Hungarian conversational model*, Annales Mathematicae et Informaticae 61 (2025), pp. 261–274, doi: 10.33039/ami.2025.10.010.