

PINNACLE: Public Engagement in AI Evaluation for Education and System Benchmarking

Maria Kasinidou¹, Jahna Otterbacher^{1,2}, Styliani Kleanthous¹, Evgenia Christoforou¹

¹Open University of Cyprus, Cyprus,

[\(maria.kasinidou,jahna.otterbacher,styliani.kleanthous,evgenia.christoforou\)@ouc.ac.cy](mailto:(maria.kasinidou,jahna.otterbacher,styliani.kleanthous,evgenia.christoforou)@ouc.ac.cy)

²CYENS Centre for Excellence, Cyprus

INTRODUCTION

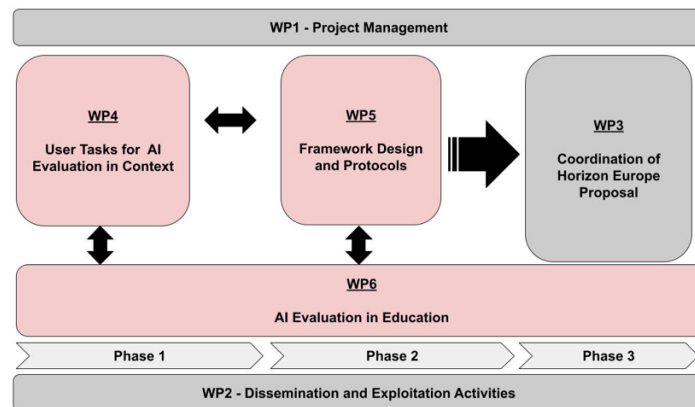
The rapid deployment of AI systems across education, work, and public services is transforming the landscape of digital skills, expanding them to include information and AI literacy [1]. Developing AI literacy is increasingly recognized as a core component of digital competence, encompassing not only understanding AI technologies but also critically evaluating AI applications and their societal impact [2], [3]. Traditionally, AI systems are evaluated through benchmarks, standardized tasks, datasets, and metrics, designed to measure system performance and properties such as safety, robustness, and fairness [4]. While these benchmarks are essential for assessing trustworthiness, they do not capture how real users perceive AI systems in practice [5]. Addressing this gap is essential not only for promoting responsible AI adoption but also for strengthening AI literacy [6]. The PINNACLE project (BRIDGE2HORIZON/0823E/0203), co-funded by the Cyprus Research and Innovation Foundation under the BRIDGE2HORIZON program, responds directly to this challenge by developing participatory AI evaluation, combining bottom-up (user-driven) and top-down (researcher-guided) approaches. Through structured tasks integrated into public AI education programs, PINNACLE enables users to engage with AI applications critically, evaluate their trustworthiness, and reflect on their own trust, thereby fostering AI literacy and providing actionable insights for both researchers and policymakers.

PROCESS AND WORK PACKAGES

The PINNACLE project is structured into six Work Packages (WPs) as depicted in Figure 1, organized into three interdependent phases: the design of user evaluation tasks, the

development of protocols and framework, and the integration of the activities into education alongside preparation for Horizon Europe proposal. WP1 and WP2 provide the backbone of the project, ensuring smooth coordination, quality assurance, and communication of the results. WP3 fulfills the BRIDGE2HORIZON objective by preparing and submitting a Horizon Europe proposal, consolidating the outputs of all activities and positioning the consortium for sustained participation in EU research. WPs 4-6 focus on the fundamental research that underlies the creation of user-centered AI evaluation methodology. WP4 conducts participatory design of evaluation tasks that are appropriate and interesting for the public, while WP5 focuses on the development of the framework, establishing the necessary protocols for collecting and aggregating data from the user-sourced evaluations, and the required analytics to produce insights. WP6 considers the integration of AI evaluation activities into our ongoing course “AI in Everyday Life” [7], [8]. Beyond the current course, WP6 designs additional educational initiatives, to foster a Participatory AI culture in the local community.

Figure 1. PINNACLE work packages and their interdependencies.



MOST RELEVANT RESULTS

Results to date

Building on the participatory evaluation framework developed in the PINNACLE project, the “AI in Everyday Life” course, offered since 2021 at the Open University of Cyprus, engaged participants in structured tasks to reflect on AI applications they use in daily life. In line with the project’s dual approach, participants completed bottom-up tasks - identifying AI-enabled applications independently, and top-down tasks - analyzing specific technologies aligned with weekly course topics. Each week, participants documented relevant applications, selected one for detailed evaluation, and assessed their trust on the application and its trustworthiness. Analysis of the collected logs shows that participants’ trust ratings were generally consistent with

their trustworthiness assessments, indicating appropriate alignment between perception and evaluated quality. Trust also proved to be dynamic, varying over time and across different AI technologies. Preliminary analysis also showed that the course and the tasks enabled participants to evaluate AI in their everyday life more critically and advance their AI literacy in general. These findings confirm that the structured, reflective tasks can effectively produce meaningful measures of perceived trust and trustworthiness in real-world contexts. Overall, these preliminary results demonstrate that the PINNACLE methodology not only supports user-centered AI evaluation but also fosters digital competence and AI literacy.

Expected Results

Building on the preliminary piloting of the evaluation tasks, PINNACLE expects to deliver additional outputs that extend the impact of these tasks. While initial work demonstrated the feasibility of embedding participatory evaluation tasks into educational settings, the project will refine these protocols and best-practice guidelines to support broader implementation across diverse classrooms and community contexts. The project will also produce expanded educational resources, including lesson plans, activity guides, and case studies, to facilitate the integration of bottom-up and top-down AI evaluation tasks into AI lifelong learning programs. Aggregated datasets from pilots across multiple initiatives will enable comparative analyses of trust, fairness, and autonomy perceptions, highlighting demographic and cultural variations. Interactive dashboards and visualization tools will make these insights accessible to educators, researchers, and policymakers.

CONCLUSION

The PINNACLE project advances AI literacy using participatory evaluation tasks with user-centered methodologies, enabling citizens not only to understand AI technologies but also to critically assess their trustworthiness. By integrating bottom-up and top-down approaches into educational programs, the project demonstrates that meaningful user perspectives can complement traditional technical benchmarks. These insights support responsible AI adoption, foster digital competence, and provide a scalable framework for embedding participatory AI evaluation into broader educational and community contexts.

ACKNOWLEDGMENT

This work has received funding from the Cyprus Research and Innovation Foundation under the BRIDGE2HORIZON program with agreement No. BRIDGE2HORIZON/0823E/0203 (PINNACLE).

REFERENCES

- [1] S. Gonzales, "AI literacy and the new digital divide - A global call for action," UNESCO, Aug. 06, 2024.
<https://www.unesco.org/en/articles/ai-literacy-and-new-digital-divide-global-call-action>
- [2] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–16, Apr. 2020, doi: <https://doi.org/10.1145/3313831.3376727>.
- [3] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. Q. Shen, "Conceptualizing AI literacy: An Exploratory Review," Computers and Education: Artificial Intelligence, vol. 2, no. 1, Nov. 2021, doi: <https://doi.org/10.1016/j.caeai.2021.100041>.
- [4] D. Schlangen, "Targeting the Benchmark: On Methodology in Current Natural Language Processing Research," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Short Papers), Aug. 2021.
- [5] T. A. Bach, A. Khan, H. Hallock, G. Beltrão, and S. Sousa, "A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective," International Journal of Human-Computer Interaction, vol. 40, no. 5, pp. 1–16, Nov. 2022, doi: <https://doi.org/10.1080/10447318.2022.2138826>.
- [6] S. Gnoth and J. Novak, "Supporting AI Literacy Through Experiential Learning: An Exploratory Study," Lecture Notes in Computer Science, In International Conference on Human-Computer Interaction, p. pp. 233–251, 2025, doi: https://doi.org/10.1007/978-3-031-93746-0_17.
- [7] M. Kasinidou, Styliani Kleanthous, and J. Otterbacher, "Artificial Intelligence in Everyday Life: Educating the Public Through an Open, Distance-learning Course," Proceedings of the 2023 on Innovation and Technology in Computer Science Education V. 1, Jun. 2023, doi: <https://doi.org/10.1145/3587102.3588784>.
- [8] M. Kasinidou, S. Kleanthous, M. Busso, M. Rodas, J. Otterbacher, and F. Giunchiglia, "Artificial Intelligence in Everyday Life 2.0: Educating University Students from Different Majors," Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1, pp. 24–30, Jul. 2024, doi: [10.1145/3649217.3653542](https://doi.org/10.1145/3649217.3653542).