

Beyond Adaptivity: A Modular NLP-Driven Framework for Dynamic Learner Profiling and Generative Educational Content

Modestino Matarazzo ^{1*}, Roberto Caldelli ¹, Barbara Martini ¹ and Filippo Sciarrone ¹

¹ Universitas Mercatorum; modestino.matarazzo@unimercatorum.it;
roberto.caldelli@unimercatorum.it; barbara.martini@unimercatorum.it;
filippo.sciarrone@unimercatorum.it

* Correspondence: modestino.matarazzo@unimercatorum.it

Abstract: Generative Artificial Intelligence (GAI) is rapidly transforming the education landscape, enabling both teachers and students to benefit from new forms of personalization. However, systems that aim to be adaptive are often tied to, and therefore limited by, predefined and static parameters, which limit their adaptability to the user. This article presents a modular framework that uses natural language processing (NLP) technologies to extract significant features from student-written contributions to build dynamic profiles that model their abilities and needs. These generated profiles are then used to guide Large Language Models (LLMs) in producing learning content that can be deemed adaptive and personalized. The proposed system integrates and aims to offer a real benefit to both students and teachers. The article will focus on the description of the designed pipeline, its applications in different educational contexts, the evaluation methodology, and, finally, the resulting ethical safeguards.

Keywords: Dynamic Learner Profiling; Generative Educational Content; NLP in Education



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, most adaptive platforms and learning analytics systems currently rely primarily on behavioral indicators, such as quiz scores, time spent on homework, or interaction patterns, to infer student needs (Kasneci et al., 2023), (Zawacki-Richter et al., 2019). While these approaches are highly effective at identifying knowledge gaps, they rarely capture the true semantic and metacognitive richness that emerges from the natural language used and produced by students in their activities. At the same time, the application of large-scale linguistic models in education has focused primarily on content delivery, such as the automatic generation of questions, explanations, or educational dialogues (Rau & Zahn, 2022), (Sharma et al., 2025).

Profiling without the generation of adaptive content risks remaining purely diagnostic, but ineffective in directly supporting students. On the other hand, content generation without profiling risks remaining superficial and never truly tailored to the student's needs (Rodríguez-Ortiz et al., 2025). To achieve a solution that combines these two aspects, we need to consider a closed-loop architecture capable of building

dynamic profiles based on students' authentic language and consistently using these profiles to guide the generation of truly personalized educational content.

The framework we propose, which we will call AdaP (Adaptive Profiler), extracts lexical, syntactic, pragmatic, affective, and argumentative characteristics. This information is combined into a dynamic student profile that evolves over time, reflecting both stable aspects (such as communication style or linguistic register) and aspects that can vary (such as motivation or emotional state). The profile constructed is then used to build prompts and guide the post-editing phase of the generative models, ensuring that everything produced is adapted to the constantly evolving needs of each learner.

2. Materials and Methods

The primary motivation for this research stems from three interconnected shortcomings that currently characterize the landscape of adaptive learning and GAI applied to education (Kasneci et al., 2023), (Zawacki-Richter et al., 2019). (i) In a learning adaptive context, most learning technologies still rely on quantitative performance indicators (Zawacki-Richter et al., 2019). These indicators, while useful for monitoring student progress, fail to capture aspects such as the semantic and meta-cognitive richness present in the natural language. (ii) Research in NLP field applied to education has demonstrated that students' written or oral contributions can reveal not only their lexical and syntactic skills, but also affective states, perceived self-efficacy, and argumentative strategies (Holstein et al., 2019). However, these linguistic cues are rarely integrated systematically into adaptive learning systems. This shortcoming represents a missed opportunity to construct student profiles in a cognitively and pedagogically meaningful way. (iii) Recent research on the use of LLM in education, particularly for content generation, has highlighted their ability to produce explanations, exercises, and didactic dialogues with surprising fluency (Kasneci et al., 2023), (Rau & Zahn, 2022). These generative systems, however, function without reference to fine-grained student profiles, producing content that may be engaging but insufficiently targeted to learners' needs. These three limitations, taken together, leads to the idea of a closed-loop system that would not only diagnose students' needs from authentic linguistic data but also act on that information by adapting, for example, the tone, complexity, and level of scaffolding of the generated materials.

AdaP was developed as a modular system in Python and consists of six main elements:

1. Data collection and preprocessing. Texts produced by students are collected through a FastAPI service. A privacy-preserving preprocessing step anonymizes personally identifiable information (PII) such as names, emails, and phone numbers using Named Entity Recognition (NER) tools (e.g., spaCy NER, Microsoft Presidio) and replaces them with hashed identifiers (Neamatullah et al., 2008), (Fung et al., 2010).
2. Linguistic feature extraction. Using spaCy (Honnibal et al., 2020), texts are analyzed to compute a FeatureVector that captures lexical richness (token count, type-token ratio, MTLT), syntactic complexity (mean sentence length, dependency distance) (Lu, 2011), pragmatic cues (hedges, questions, first-person pronouns), affective characteristics (sentiment polarity and emotion categories, using NRC Emotion Lexicon), readability indices

- (Flesch–Kincaid, Gunning Fog) (Kincaid et al., 1975), and heuristic argument quality.
3. Dynamic profile construction. Features are aggregated into a Learner Profile object that stores both stable attributes (tone, register) and dynamic ones (CEFR level, affective state, motivation index). Profile updates use an Exponential Moving Average (EMA, $\alpha = 0.25$) to capture both long-term tendencies and short-term fluctuations (Hunter, 1986). CEFR classification is estimated using supervised approaches aligned with computational CEFR prediction studies (Vajjala & Rama, 2018).
 4. Rule-based generation control. A transparent policy engine maps learner profile attributes to generation parameters, such as tone (supportive, neutral), register (plain, semi-formal), readability grade, number of scaffolding steps, and inclusion of examples. These rules are interpretable and align with best practices for explainable AI in education (Holstein et al., 2019).
 5. Conditional prompt construction and generation. A structured prompt combines system directive, a snapshot of the learner profile, and the pedagogical task (explanation, exercise, or feedback). The system can operate in two modes: (i) mock mode for offline testing, and (ii) LLM mode using external APIs such as GPT-4 (Achiam et al., 2023).
 6. Post-editing and traceability. The generated content undergoes lightweight post-editing to ensure readability simplification, lexical detoxification (Xu et al., 2020), and formatting consistency. The system returns both the final output and a trace containing the learner profile snapshot, policy decisions, raw LLM output, and the post-edited version, ensuring auditability and transparency (Floridi et al., 2018).

To analyze the pedagogical effectiveness of AdaP, a mixed-methods evaluation is proposed (Creswell & Creswell, 2017): (i) Knowledge improvements will be measured with pre and post-intervention tests, comparing experimental and control groups. (ii) Interaction logs will be analyzed (time on homework, frequency of interactions, voluntary use of adaptive explanations). (iii) Questionnaires and interviews will assess students' subjective perceptions of the clarity, relevance, and motivational impact of the generated content. (iv) Experts (teachers) will evaluate the materials produced based on linguistic correctness, conceptual accuracy, and pedagogical adequacy (Shute, 2008). The combination of a modular and traceable architecture with empirical validation allows the framework to be examined both as a technical artifact and as a pedagogical intervention.

3. Results

In this section, we describe a simulated experiment aimed at modeling both the functioning and expected impact of the AdaP framework. The simulation reproduced a six-week university course, involving 180 hypothetical students, randomly assigned to three experimental conditions: (1) Control, with static materials without AI support ($n = 60$); (2) Generic-LLM, in which content is generated by a large language model without personalization ($n = 60$); (3) Profiled-LLM, in which content generation occurs via AdaP, conditioned on dynamic student profiles ($n = 60$).

To emulate the main data collected in real-world experimental settings, synthetic datasets were created, which can be summarized as follows: learning outcomes (pre- and post-test scores on 30 items), engagement levels (time spent on the platform per week), perception of personalization (Likert scale 1–7), content quality, as assessed by expert users (clarity, accuracy, pedagogical value; scale 1–5), and safety and transparency indicators (frequency of reported outputs and use of dashboards by teachers). Numerical values were generated to maintain internal consistency and empirical plausibility, allowing for the application of standard statistical analyses (ANCOVA, ANOVA, Kruskal–Wallis, MANOVA). Subgroup analyses were also simulated based on CEFR language proficiency level (A2/B1 vs. B2+) and ablation studies were conducted to assess the contribution of the core AdaP modules. Table 1 reports the simulated mean pre-test, post-test, and gain scores (Mean $\pm$ SD; maximum = 30). Results show progressive improvement across all conditions, with the most marked increase in the Profiled-LLM group. An ANCOVA analysis, using the post-test as the dependent variable and the pre-test as a covariate, revealed a significant treatment effect, $F(2, 176) = 10.79$, $p < .001$, $\eta^2_p = 0.11$. Pairwise comparisons confirmed that Profiled-LLM outperforms both the Control ($p < .001$, $d = 0.98$) and Generic-LLM ($p = .016$, $d = 0.54$) groups, while Generic-LLM still outperforms the Control ($p = .041$, $d = 0.43$).

Table 1. Simulated pre/post performance and gains (Mean $\pm$ SD; Max=30).

Condition	Pre	Post	Gain Δ
Control	16.2 $\pm$ 4.3	19.4 $\pm$ 4.6	3.2 $\pm$ 4.5
Generic-LLM	16.1 $\pm$ 4.2	21.2 $\pm$ 4.3	5.1 $\pm$ 4.2
Profiled-LLM	16.3 $\pm$ 4.1	23.6 $\pm$ 4.0	7.3 $\pm$ 4.0

Behavioral indicators show increased engagement among students supported by adaptive systems. The median weekly time on the platform was 54 minutes for the control group, 62 for the Generic-LLM, and 71 for the Profiled-LLM; the Kruskal–Wallis test confirmed significant differences ($H(2) = 12.3$, $p = .002$).

Perceived personalization ratings (scale 1–7) also followed the same trend: Control = 3.08 ± 1.04 , Generic-LLM = 4.22 ± 1.09 , Profiled-LLM = 5.55 ± 1.03 ; the ANOVA revealed a robust effect, $F(2, 177) = 33.2$, $p < .001$, $\eta^2_p = 0.27$, with all pairwise differences significant. Expert assessments of the generated content gave the Profiled-LLM group the highest scores in terms of clarity, accuracy, and pedagogical value, with excellent inter-rater reliability ($ICC = 0.84$). Analysis by CEFR proficiency levels showed that students with lower proficiency (A2/B1) benefited the most from the Profiled-LLM condition, with simulated mean gains of 8.6 ± 4.0 points compared to 5.7 ± 4.3 for the Generic-LLM. Higher-level students (B2+) also showed improvements, albeit by a smaller margin. These results confirm AdaP's ability to adapt the intensity of scaffolding based on individual needs. To assess the contribution of individual modules, ablative tests were simulated by progressively eliminating the post-editing and policy engine components. As reported in Table 2, both removals reduced the clarity and readability of the generated texts, confirming the functional relevance of these modules in the production of pedagogically appropriate content.

Table 2. Simulated ablation effects within Profiled-LLM

Variant	Avg sentence length	Readability grade	Clarity (1–5)
Full model	21.4 words	6.8	4.42 ± 0.48
No post-edit	26.8 words	7.6	4.10 ± 0.52
No policy (fixed)	23.9 words	7.2	4.21 ± 0.50

Automatic filters flagged less than 1% of the generated outputs across the entire dataset we simulated, with no significant issues identified after manual review. Instructor surveys indicated that AdaP's transparent tracking system improved understanding of student needs (78%) and confidence in generative decisions (71%). Overall, the simulation study suggests that the Profiled-LLM configuration consistently outperforms both the control and Generic-LLM groups in terms of learning, engagement, and perceived personalization. Ablation and security analyses further show that the post-editing, policy control, and transparency components contribute substantially to the framework's educational reliability. Although the data are simulated, the structure and interpretation closely mirror the procedure envisioned for future empirical validation, thus offering a consistent preview of AdaP's expected pedagogical impact.

4. Discussion

The results of our study reinforce the idea that the shift from traditional, rule-based profiling to dynamic personalization could mark a significant paradigm shift in educational technology. Adaptive systems currently rely primarily on performance metrics, often neglecting contextual and linguistic aspects of learning. In contrast, approaches such as AdaP demonstrate that integrating linguistic, affective, and pragmatic dimensions allows for a more complete representation of students' cognitive/emotional states (Winne, 2017), (Lan et al., 2024).

Recent advances in LLMs have further expanded the opportunities for adaptive tutoring, with personalized feedback and contextualized explanations (Kasneci et al., 2023), (Sharma et al., 2025). This study highlights, however, that without the explicit integration of student modeling mechanisms, these systems risk producing fluid but generic outputs, lacking true personalization. AdaP helps fill this gap by basing generation on dynamically updated profiles; It also ensures that generative processes are aligned with pedagogical goals, rather than superficial personalization.

A key dimension emerging from this work concerns transparency and understandability. In education, automation cannot replace the need for interpretability. Teachers and students must understand not only the results, but also the underlying motivations (Holstein et al., 2019). The inclusion of a rule-based decision-making layer and a traceable generation pipeline supports this need, providing an example of explainable AI that can enhance teacher autonomy and accountability (Holstein et al., 2019), (Hu & Wang, 2024), (Luckin & Holmes, 2016). This transparent approach aligns with the movement toward ethical, participatory, and human-centered AI in learning environments (Floridi et al., 2018). The ethical implications of using GAI in education remain critical. The concepts of fairness, bias mitigation, data privacy, and accountability remain fundamental prerequisites for trustworthy AI-mediated pedagogy (Floridi et al., 2018), (Wu et al., 2025). The expansion of adaptive systems across

diverse cultural and linguistic contexts also raises issues of inclusivity and equity. Another important aspect concerns student autonomy. While personalization can improve engagement and learning outcomes, it is necessary to maintain an appropriate balance between adaptivity and student autonomy (Mohajan, 2025). Excessive adaptation can limit exploration and reduce opportunities for students' productive engagement, which are essential for deep learning. Systems like AdaP should be designed as learning support mediators but must also preserve student autonomy and self-regulation. The sustainability of GAI-based personalization ultimately depends not only on technical performance but also on its institutional and cultural acceptance (Maity & Derooy, 2024). Teacher trust, curricular alignment, and ethical compliance are crucial for long-term adoption. The modular and traceable architecture proposed by AdaP represents a first step in this direction, but full integration into real-world educational practice will require iterative testing, teacher engagement, and systematic evaluation.

We can therefore say that, overall, this discussion emphasizes that the future of adaptive learning depends on the convergence of three interdependent principles: (i) dynamic, language-driven learner modeling; (ii) transparent and ethically grounded generative mechanisms; and (iii) sustainable integration into authentic educational practices. Only through this alignment can the use of GAI evolve from a pure technological innovation to a significant pedagogical partner.

5. Conclusions

This study introduced AdaP, a modular framework that integrates NLP, dynamic learner profiling, and rule-based generative adaptation. AdaP bridges the gap between diagnostic learner analytics and generative AI systems that often lack genuine personalization (Kasneci et al., 2023). By transforming learner texts into multidimensional profiles, capturing lexical, syntactic, pragmatic, and affective dimensions, the framework conditions large language models to produce tailored explanations, exercises, and feedback. Simulated evaluations indicate potential gains in learning outcomes and engagement, particularly for low-proficiency learners (Shute, 2008). AdaP contributes to research by demonstrating (1) the feasibility of building meaningful learner profiles from authentic language (Holstein et al., 2019); (2) the interpretability of profile-guided content generation (Luckin & Holmes, 2016); and (3) the ethical integration of anonymization and teacher-facing transparency (Floridi et al., 2018). Limitations include the heuristic nature of current feature extraction and the need for validation with real learners, improved CEFR classifiers (Vajjala & Rama, 2018), and affective modeling. In summary, AdaP supports a closed-loop approach that connects learner modeling with real-time generative support, advancing personalization toward transparency, fairness, and pedagogical reliability (Zawacki-Richter et al., 2019). Future work will focus on the empirical validation of the AdaP framework through pilot studies with real learners, assessing its effectiveness in different educational contexts and refining its adaptive mechanisms based on observed learning outcomes.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). GPT-4 technical report. arXiv:2303.08774. <https://arxiv.org/abs/2303.08774>

- Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., ... & Wu, Y. (2023). PaLM 2 technical report. arXiv:2305.10403. <https://arxiv.org/abs/2305.10403>
- Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. SAGE.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Fung, B. C., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4), 1–53.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3290605.3300830>
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in Python.
- Hu, S., & Wang, X. (2024, August). FOKE: A personalized and explainable education framework integrating foundation models, knowledge graphs, and prompt engineering. In *China National Conference on Big Data and Social Computing* (pp. 399–411). Springer.
- Hunter, J. S. (1986). The exponentially weighted moving average. *Journal of Quality Technology*, 18(4), 203–210.
- Kasneci, E., Sessler, K., Betschart, M., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel (RBR-75-8).
- Lan, Y., Li, X., Du, H., Lu, X., Gao, M., Qian, W., & Zhou, A. (2024). Survey of natural language processing for education: Taxonomy, systematic review, and future trends. arXiv:2401.07518.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62.
- Luckin, R., & Holmes, W. (2016). Intelligence unleashed: An argument for AI in education.

- Maity, S., & Deroy, A. (2024). Generative AI and its impact on personalized intelligent tutoring systems. *arXiv:2410.10650*.
- Mohajan, H. K. (2025). Artificial intelligence: Prospects and challenges in future progression. *Art and Society*, 4(7), 38–50.
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436–465. <https://doi.org/10.1111/j.1467-8640.2012.00460.x>
- Neamatullah, I., Douglass, M. M., Lehman, L. W. H., Reisner, A., Villarroel, M., Long, W. J., ... & Clifford, G. D. (2008). Automated de-identification of free-text medical records. *BMC Medical Informatics and Decision Making*, 8(1), 32. <https://doi.org/10.1186/1472-6947-8-32>
- Rau, M. A., & Zahn, M. (2022). Nonverbal collaboration on perceptual learning activities with chemistry visualizations. In M. M. Rodrigo et al. (Eds.), *Artificial Intelligence in Education* (Vol. 13356). Springer. https://doi.org/10.1007/978-3-031-11647-6_41
- Rodríguez-Ortiz, M. Á., Santana-Mancilla, P. C., & Anido-Rifón, L. E. (2025). Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges. *Applied Sciences*, 15(15), 8679. <https://doi.org/10.3390/app15158679>
- Sharma, S., Mittal, P., Kumar, M., & Bhardwaj, V. (2025). The role of large language models in personalized learning: A systematic review of educational impact. *Discover Sustainability*, 6(1), 1–24.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Vajjala, S., & Rama, T. (2018). Experiments with universal CEFR classification. *arXiv:1804.06636*.
- Wang, X. J., Lee, C. P., & Mutlu, B. (2025, April). LearnMate: Enhancing online education with LLM-powered personalized learning plans and support. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1–10).
- Winne, P. H. (2017). Learning analytics for self-regulated learning. In *Handbook of Learning Analytics* (pp. 241–249). Society for Learning Analytics Research. <https://doi.org/10.18608/hla17.020>
- Wu, F., Dang, Y., & Li, M. (2025). A systematic review of responses, attitudes, and utilization behaviors on generative AI for teaching and learning in higher education. *Behavioral Sciences*, 15(4), 467.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>