DAVE: INTEGRATING PERSONALIZED AI TUTORING FOR ENGAGED LEARNING

Mr. Benjamin Joseph Spiteri
Prof. Alexiei Dingli

Traditional educational systems often struggle to cater to the diverse learning needs and requirements of individual students. This one-size-fits-all approach can hinder unique development, disengage learners, and exacerbate existing inequalities within the classroom. This paper presents DAVE (Digital Autonomous Virtual Educator), a personalized AI tutoring system designed to seamlessly integrate within the engaging FAIE learning application used in Maltese primary schools. DAVE leverages the power of Large Language Models (LLMs) and an advanced automated prompt engine to provide tailored support and adaptive feedback in mathematics. By analyzing real-time academic data from FAIE, DAVE adapts learning materials and offers personalized explanations and exercises, creating a dynamic and responsive learning environment. Through evaluations involving students in educational situations comparing DAVE with a commercially available LLM-powered chatbot, results demonstrated DAVE's significant impact on student learning outcomes, particularly for students with lower mathematical proficiency. Students utilizing DAVE achieved improved results on mathematics worksheets and reported higher user satisfaction, emphasizing DAVE's helpfulness, clear explanations, and personalized support. These findings underscore DAVE's potential to bridge the learning gap, promote educational equity, and foster a more engaging and effective learning experience for all students. The seamless integration of DAVE within the FAIE platform minimizes disruption to existing workflows and maximizes student engagement. This research helps to highlight the potential of AI-powered educational tools to transform the learning landscape, offering personalized support, mitigating misinformation, and empowering students to achieve greater academic success.

Keywords: Personalized learning, LLMs, Prompt Engineering, NLP

DOI: **https://doi.org/10.48341/7q0t-wj47**

Introduction

Personalized learning is an educational approach tailored to the individual needs and learning styles of students. This approach has emerged as a critical factor in maximizing student engagement and academic success. Through recognizing that each student learns at their own pace and using different modalities, personalizing learning pathways offers a compelling alternative to traditional one-size-fits-all models. By providing customized learning materials, adaptive feedback, and individualized support, personalized learning empowers students to take ownership of their educational journey, fostering deeper understanding, increased motivation, and improved learning outcomes. However, implementing personalized learning at scale presents significant challenges, particularly in primary education where teachers often face large class sizes and limited resources. Effectively addressing this challenge requires innovative solutions that leverage the power of technology while maintaining the crucial role of human educators in fostering a positive and supportive learning environment. The thoughtful and holistic integration of technology within educational frameworks is essential to ensure that these tools enhance, rather than hinder, the learning process.

In the context of Maltese primary schools, the FAIE learning application stands as a prime example of technology's potential to create engaging and interactive learning experiences. Developed as part of the EducationAI project, FAIE provides students in grades 4-6 with access to a wide range of educational resources, interactive exercises, and progress tracking tools. The application's user-friendly interface and gamified elements encourage active participation and foster a positive attitude towards learning. However, while FAIE provides a valuable platform for engaging students, it lacks the capacity to offer truly personalized support and guidance.

To further enhance the capabilities of the FAIE application and prove the viability of LLMs for education, this research introduces DAVE, an AI-powered tutoring system designed to seamlessly integrate with the FAIE platform. DAVE acts as a personalized AI companion for each student, providing tailored support, adaptive feedback, and individualized learning pathways in mathematics. By leveraging the capabilities of LLMs, prompt engineering and introducing a novel verification framework, DAVE offers a dynamic and responsive learning experience that adapts to each student's unique needs and learning style. DAVE's integration within FAIE aims to bridge the gap between engaging educational content and personalized learning support, creating a synergistic environment that maximizes student engagement and promotes educational equity.

The development of DAVE was guided by the recognition that technology should serve as a supporting tool for educators and students, not as a replacement for the essential human element in education. DAVE's primary role is to enhance, not replace, the teacher's role by providing students with an additional resource for individualized support and guidance. This allows teachers to focus on higher-level tasks such as facilitating discussions, fostering critical thinking, and addressing the individual needs of students who require more direct intervention. DAVE's integration within FAIE aims to create a collaborative learning environment where technology and human interaction work in concert to empower students to reach their full potential.

Literature Review

The need for modernizing the educational approach has continued to be established time and time again, highlighting the importance of promoting individuality and promoting the unique talents of each student. This approach focuses on tailoring to the individual needs and learning styles of students. Personalized learning has gained significant traction in recent years as a key strategy for enhancing educational outcomes. Unlike the traditional one-size-fits-all approaches, personalized learning recognizes that each student learns at their own pace and through different modalities (Williams, 2015). It empowers students to take ownership of their learning journey by providing customized learning materials, adaptive feedback, and individualized support (Bernacki, 2021). This learner-centric approach fosters deeper understanding, increased motivation, and improved learning outcomes across various subjects and age groups (Pane, 2015). Furthermore, personalized learning promotes the development of essential 21st-century skills, such as self-directed learning, metacognition, and adaptability (Bray, 2015), preparing students for success in a rapidly evolving world. By catering to diverse learning styles such as visual, auditory, kinesthetic, while also allowing students to progress at their own pace, personalized learning creates a more inclusive and equitable learning environment (Onyishi, 2020). This approach aligns with key educational philosophies like constructivism, which emphasizes the active role of learners in constructing knowledge (Tan, 2019), and humanism, which underscores the importance of nurturing each student's unique talents and interests (Aung, 2020).

Intelligent Tutoring Systems (ITS) have emerged as a promising technological tool for implementing personalized learning. ITS are computer-based software systems designed to provide students with personalized instruction and support by dynamically adapting educational materials to their unique learning styles (Graesser, 2012). These systems act as a bridge between educators and students, offering a digital learning environment that complements classroom instruction (Ma, 2014). Early ITS relied on rule-based systems and decision trees, limiting their ability to recognize individual learning patterns and provide truly personalized feedback (Nwana, 1990). However, with the advent of AI, particularly machine learning and knowledge representation, ITS have undergone a transformative evolution. AI-powered ITS leverage sophisticated algorithms to analyze student data, discern intricate learning patterns, and adjust instruction in real-time to optimize the learning experience (Alam, 2023). These systems can provide personalized feedback, tailor learning pathways, and offer targeted support based on student performance data (Lin, 2023). Research has consistently demonstrated the efficacy of AI-powered ITS in enhancing student learning outcomes across various subjects (VanLehn, 2006).

Despite the advancements in AI-powered ITS, traditional systems face limitations in handling complex, open-ended tasks and often rely on pre-defined content, restricting their adaptability and responsiveness to individual student needs (Feng, 2021). The emergence of generative AI, particularly Large Language Models (LLMs), offers a potential solution to these limitations. LLMs, trained on massive datasets, possess remarkable capabilities in understanding and generating human-like text (Zhao, 2023). In the context of education, LLMs can create personalized learning materials, provide adaptive feedback, and engage in natural language dialogue with students, offering a more dynamic and interactive learning experience (Latham, 2022). This personalized and interactive approach has the potential to significantly enhance student engagement and motivation, fostering a positive and productive learning environment (Ji, 2022). DAVE leverages the power of LLMs to provide tailored support and guidance to students, adapting to their individual learning needs and promoting a more self-directed approach to learning (Hemachandran, 2022).

Generative AI, powered by LLMs, offers a transformative approach to education by enabling the creation of personalized learning experiences. These models can generate customized learning materials, such as tailored explanations, practice exercises, and study plans (Okonkwo, 2021). Furthermore, LLMs can provide adaptive feedback on student work, addressing individual misconceptions and offering targeted guidance for improvement (Ji, 2022). By engaging in natural language dialogue, LLMs can also offer on-demand support, answering student questions and providing clarification on complex concepts (Stamper, 2024). This personalized and interactive approach has the potential to significantly enhance student engagement and motivation, fostering a positive and productive learning environment. However, it is crucial to acknowledge the limitations of LLMs, such as their susceptibility to generating inaccurate or misleading information know as hallucinations (Lee, 2024) and potential biases embedded within their training data. Carefull considerations need to be made to ensure that generated responses are verified in order to mitigate the potential risk of hallucinations.

The implementation of AI-powered educational tools requires careful consideration of ethical implications and effective implementation strategies (Holmes, 2022). Data privacy is paramount, and robust data protection measures, along with transparent data usage policies, are essential for safeguarding student information (Kumar, 2024). Addressing algorithmic bias is crucial to ensure fairness and equity in educational opportunities (Tyser, 2024). Furthermore, maintaining the human element in education, including the student-teacher relationship, is vital for fostering a positive and supportive learning environment (Zheng, 2022). Effective implementation strategies involve providing comprehensive training for both students and educators on the use of AI-powered tools, addressing both technical aspects and ethical considerations. Furthermore, ensuring access to appropriate devices and reliable internet connectivity is crucial for equitable access to these resources (De La Higuera, 2019). By addressing these ethical considerations and implementing thoughtful strategies for content verification, AI-powered educational tools can be effectively integrated into the learning landscape, maximizing their potential to enhance student learning while minimizing potential risks.

Methodology

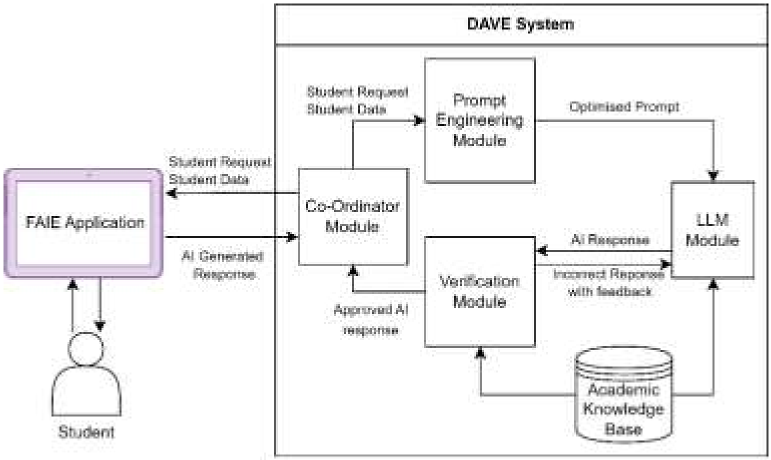

Figure 1. DAVE System Architecture (Source: Self).

DAVE is designed as a modular system to provide personalized learning support within the FAIE application. This modular architecture promotes flexibility, scalability, and maintainability, enabling future expansion and adaptation to evolve educational needs. The system comprises four interconnected modules, each with distinct functionalities: the Coordinator Module, the Prompt Engineering Module, the LLM Module, and the Verification Module. These modules work in unison to provide a seamless and personalized learning experience within the FAIE environment through a chatbot portal. The system leverages real-time academic data from FAIE to personalize the learning experience and employs the robust FORT Verification framework to ensure the safety and reliability of LLM-generated responses.

The Prompt Engineering Module plays a crucial role in personalizing the interaction between the student and the LLM. Guided by user-centered design principles, this module transforms the student's raw input into an optimized prompt tailored to their individual learning needs and the specific context within FAIE. This optimization process leverages a combination of prompt engineering techniques, each chosen for its effectiveness in enhancing the LLM's performance and creating a more engaging user experience. Role assignment establishes a helpful and supportive persona for DAVE, fostering a positive learning environment. Template prompting ensures consistent and structured input to the LLM, simplifying interactions for young learners and maximizing the clarity of requests. Deep context integration leverages real-time data from FAIE, including the student's current workflow, proficiency levels, and specific learning objectives. This contextual information allows DAVE to provide more relevant and targeted support. Through the real time adjustment of explanations and examples, DAVE ensures that any student struggling with a given concept is provided with tailored assistance supporting them to overcome their limitations. Few-shot prompting provides the LLM with examples of desired interactions, enhancing its ability to understand and respond appropriately to student queries. Chain-of-Thought prompting encourages step-by-step reasoning, guiding the LLM to break down complex problems into smaller, more manageable steps and promoting deeper understanding. Additionally Chain-of-Thought helps to promote educational scaffolding, helping not only the LLM but also the students understand the reasoning patterns required for problem solving. This comprehensive approach to prompt engineering ensures that the LLM receives optimized input, leading to more accurate, relevant, and pedagogically sound responses.

The LLM Module is at the center of DAVE, responsible for generating personalized responses to student queries. DAVE utilizes a fine-tuned Gemini 1.0 Pro model, accessed through the Google Vertex AI platform. This model was chosen for its advanced capabilities in natural language understanding, generation, and reasoning, its large context window, and the host platform's flexibility for fine-tuning and integration with Retrieval Augmented Generation (RAG). Crucially, each student is assigned a unique LLM instance, enabling truly personalized interactions and allowing the model to adapt to individual learning patterns over time. This personalized approach ensures that each student's LLM instance evolves based on their specific needs and interactions with DAVE. By maintaining a continuous chat history within each student's LLM instance, DAVE can track their progress, identify areas where they require additional support, and provide a more cohesive and personalized learning experience. The fine-tuning process utilizes curated mathematical datasets (GSM8K and Orca Math Word Problems 200K) to optimize the LLM's performance in the specific context of primary school mathematics. This specialization ensures that DAVE's responses are aligned with the curriculum and tailored to the students' learning objectives within FAIE.

The Verification Module implements the FORT Verification framework, a multi-stage process designed to ensure the accuracy, relevance, and safety of DAVE's responses. This framework addresses the inherent limitations of LLMs, particularly their susceptibility to generating hallucinations and misinformation.

FORT Verification incorporates twin model analysis, where a duplicate LLM instance evaluates the generated response for logical consistency and relevance. RAG, utilizing the GSM8K dataset as a knowledge base, ensures that DAVE's responses are grounded in factual mathematical information. Online search verification cross-references the response with trusted online sources, providing an additional layer of fact-checking. Finally, a feedback loop mechanism provides corrective feedback to the LLM, refining its performance over time and enhancing the quality of future responses. This robust verification process acts as a critical safeguard, ensuring that students receive accurate, relevant, and trustworthy information, promoting a safe and productive learning environment.

DAVE's seamless integration within the FAIE platform is a key design consideration. By embedding DAVE directly within FAIE, students can access personalized support without disrupting their existing workflows. This integration leverages FAIE's engaging interface and gamified elements while adding the power of personalized AI tutoring, creating a synergistic learning experience. This approach minimizes the learning curve for students and teachers, promoting adoption and maximizing the impact of DAVE within the classroom. To achieve this integration, a dedicated chatbot interface was created within FAIE that allows students to interact with DAVE without having to disrupt any current activity they are completing in the main FAIE interface. From a programmatical approach, FAIE communicated with DAVE through a dedicated API, with safeguards implemented to secure data transfer.

Evaluation

To evaluate DAVE's effectiveness in a real-world setting, a study was conducted with 65 sixth-grade students across two Maltese primary schools. This study employed a mixed-methods approach, combining quantitative analysis of student performance on mathematics worksheets and qualitative feedback on user experience. The evaluation compared DAVE with ChatGPT-4, a commercially available LLM-powered chatbot, to assess the added value of DAVE's personalized features and integration within the FAIE learning application. Two specifically designed mathematics worksheets, incorporating problems of increased difficulty compared to the students' typical coursework, were used to encourage reliance on the AI assistants. Students completed one worksheet with DAVE and the other with ChatGPT-4, with the order counterbalanced to mitigate order effects. Following the worksheet activities, students completed a user experience questionnaire providing feedback on both AI assistants. Post worksheet completion, students where asked to complete a short evaluation survey containing three parts. The first two parts of the survey required students to evaluate their experience using the two chatbots, raking them based on 6 performance metrics (relevance of responses, helpfulness of responses, ability to understand, directness of responses, situational awareness, and overall experience ranking). The third part of the survey required students to provide a self-evaluation ranking their own mathematical capabilities, and their experience using AI-chatbots.

Quantitative Results: Worksheet Performance

The quantitative analysis focused on student performance on the mathematics worksheets. Overall, students achieved significantly higher scores when using DAVE (M = 77%) compared to ChatGPT-4 (M = 45%). This difference of 32% was determined to be statistically significant (Wilcoxon Signed-Rank test, $W = 164$, $p < .01$, $r = 0.64$), indicating that DAVE significantly enhanced students' ability to solve mathematical problems.

Table 1. Average User Experience Rating and Worksheet Score by self-ability Rating (Source: Self)

	GPT4 Overall	DAVE Overall	GPT4 Worksheet	DAVE worksheet	Difference
Average ability 0-3	8	8	17%	79%	**63%**
Average ability 4-7	7	8	52%	73%	**22%**
Average ability 8-10	8	9	48%	78%	**31%**

To further investigate DAVE's impact on students with varying mathematical proficiencies, participants were grouped based on their self-rated mathematical abilities (weak, moderate, strong). The analysis revealed a striking difference in performance gains across proficiency levels. Students in the "weak" group showed the most substantial improvement, achieving an average score of 79% with DAVE compared to just 17% with ChatGPT-4, a remarkable 62% difference. The "moderate" group also benefited from using DAVE, scoring an average of 73% compared to 52% with ChatGPT-4 (a 21% improvement). While the "strong" group demonstrated high performance with both systems, they still achieved a noticeable 30% improvement with DAVE (78% vs. 48% with ChatGPT-4). These differences help in demonstrating DAVE's efficacy in supporting learners across a range of mathematical abilities, with the most significant impact observed for students who typically struggle with mathematics.

Qualitative Results: User Experience Feedback

The qualitative feedback collected through the user experience questionnaires provided valuable insights into students' perceptions of DAVE and ChatGPT-4. Overall, students rated DAVE higher across all six evaluation categories. The average user experience rating for DAVE was 8.3 out of 10, compared to 7.1 for ChatGPT-4. Students frequently praised DAVE's personalized explanations and step-by-step guidance, with one student stating, "I like DAVE because he is responsible and talks to me like he is my friend." This comment reflects DAVE's design goal of creating a supportive and approachable learning companion. Several students highlighted DAVE's seamless integration within FAIE, outlining the importance of the ease of transitioning between the familiar learning application and the personalized tutoring environment. While generally positive, some feedback pointed to the possible excessive length of DAVE's explanations as an area for potential improvement.

About the Author

Mr. Benjamin Spiteri is a researcher and recent graduate from the University of Malta, holding a Master of Science in Artificial Intelligence. His research interests include LLMs, Computer Vision, and Game Development. As a full time researcher at the University of Malta, he has contributions on projects exploring AI in education, robotics, medical applications of AI, and the development of advanced game-AI.

Prof Alexiei Dingli is an AI expert and Professor at the University of Malta. With over 20 years of experience in the field, he has helped numerous companies successfully implement AI solutions. His work has been recognized as world-

class by international experts, and he has received numerous awards from organizations such as the European Space Agency, the World Intellectual Property Organization, and the United Nations. In addition to his considerable peer-reviewed publications, he is also a member of the Malta.AI task force, working to position Malta as a global leader in AI.

References

Williams, S. L. P. (2015). Personalized learning: A transformative vision. Phi Delta Kappan, 97(1), 56–62.

Onyishi, C. N., & Sefotho, M. M. (2020). Teachers' perspectives on the use of differentiated instruction in inclusive classrooms: Implication for teacher education. International Journal of Higher Education, 9(2), 136–150. https://doi.org/10.5430/ijhe.v9n2p136

Tan, C., & Ng, C. S. (2019). Constructivism in education. In Oxford research encyclopedia of education. https://doi.org/10.1093/acrefore/9780190264093.013.94

Aung, Y. M. (2020). Humanism and education. International Journal of Advanced Research in Science, Engineering and Technology, 7, 13555–13561.

Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose? Educational Psychology Review, 33(4), 1675–1715. https://doi.org/10.1007/s10648-020-09591-8

Bray, B., & McClaskey, K. (2015). Make learning personal: The what, who, wow, where, and why. Corwin Press.

Pane, J. F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2015). Continued progress: Promising evidence on personalized learning. RAND Corporation.

Zheng, F. (2022). Fostering students' well-being: The mediating role of teacher interpersonal behavior and student–teacher relationships. Frontiers in Psychology, 12, Article 796728. https://doi.org/10.3389/fpsyg.2021.796728

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology, 106(4), 901–918. https://doi.org/10.1037/a0037123

Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. American Psychological Association. https://doi.org/10.1037/13276-000

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. Artificial Intelligence Review, 4(4), 251–277. https://doi.org/10.1007/BF00148175

Alam, A. (2023). Harnessing the power of AI to create intelligent tutoring systems for enhanced classroom experience and improved learning outcomes. In Intelligent communication technologies and virtual mobile networks (pp. 571–591). Springer.

VanLehn, K. (2006). The relative effectiveness of human and intelligent tutoring systems. Educational Psychologist, 40(1), 29–45. https://doi.org/10.1207/s15326985ep4001_4

Lin, C.-C., Huang, A. Y., & Lu, O. H. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. Smart Learning Environments, 10(1), Article 41. https://doi.org/10.1186/s40561-023-00238-8

Feng, S., Magana, A. J., & Kao, D. (2021). A systematic review of literature on the effectiveness of intelligent tutoring systems in STEM. In 2021 IEEE Frontiers in Education Conference (FIE) (pp. 1–9). IEEE. https://doi.org/10.1109/FIE49874.2021.9642544

Latham, A. (2022). Conversational intelligent tutoring systems: The state of the art. In Women in computational intelligence (pp. 77–101). Springer.

Ji, S., & Yuan, T. (2022). Conversational intelligent tutoring systems for online learning: What do students and tutors say? In 2022 IEEE Global Engineering Education Conference (EDUCON) (pp. 292–298). IEEE. https://doi.org/10.1109/EDUCON54143.2022.9730707

Hemachandran, K. (2022). Artificial intelligence: A universal virtual tool to augment tutoring in higher education. Computational Intelligence and Neuroscience, 2022, Article 7292261. https://doi.org/10.1155/2022/7292261

Stamper, J., Xiao, R., & Hou, X. (2024). Enhancing LLM-based feedback: Insights from intelligent tutoring systems and the learning sciences. In International Conference on Artificial Intelligence in Education (pp. 32–43). Springer.

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. Computers and Education: Artificial Intelligence, 2, Article 100033. https://doi.org/10.1016/j.caeai.2021.100033

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., & Dong, Z. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. British Journal of Educational Technology, 55(6), 1982–2002. https://doi.org/10.1111/bjet.13375

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(3), 542–570. https://doi.org/10.1111/ejed.12540

Kumar, A., Singh, S., Murty, S. V., & Ragupathy, S. (2024). The ethics of interaction: Mitigating security threats in LLMs. arXiv preprint arXiv:2401.12273.

Tyser, K., Segev, B., Longhitano, G., Zhang, X.-Y., Meeks, Z., Lee, J., Garg, U., & Belsten, N. (2024). AI-driven review systems: Evaluating LLMs in scalable and bias-aware academic reviews. arXiv preprint arXiv:2408.10365.

De La Higuera, C. (2019). A report about education, training teachers and learning artificial intelligence: Overview of key issues. Education and Computer Sciences, 1(1), 1–11.