# Bidirectional Feedback-Based Personalization of Learning using Multi-tier AI: A Real-World Assessment of its Efficacy in Classrooms

**Saptarshi Basu, Jayson Brown, Cherie Lum, Junsoo Park, Ashok K. Goel**

Georgia Institute of Technology
sbasu7@gatech.edu, jbrown762@gatech.edu, clum7@gatech.edu, jpark3232@gatech.edu, ashok.goel@cc.gatech.edu

**Abstract**

Jill Watson is an example of an intelligent conversational AI Teaching Assistant that has been deployed across 24 class sections in different institutions, with 1102 unique student participants and over 17000 questions from Fall 2023 to present. Jill Watson's RAG-based architecture built around OpenAI ChatGPT and user study results addresses some of the concerns related to domain knowledge, deployment, and data collection in online classrooms. In this work, a 3-tiered framework for personalization in online education using AI tools to enable a human-AI personalization loop, grounded in real-world human feedback is proposed.

## Introduction

Jill Watson (Kakar et al. 2024) is an example of an intelligent conversational AI Teaching Assistant that has been deployed across 24 class sections in different institutions, with 1102 unique student participants and over 17000 questions from Fall 2023 to present. Jill Watson RAG-based architecture (Kakar et al. 2024) built around OpenAI ChatGPT and user study results (Maiti and Goel 2025) addresses some of the concerns related to domain knowledge, deployment, and data collection in online classrooms. Building on some of the personalization opportunities highlighted in the literature (Chen et al. 2024; Wei et al. 2025), this paper proposes a three-tiered framework for personalization in AI tools in education, grounded real-world human feedback. The framework encompasses personalization at three levels: adaptation to course-specific content, AI-driven customization of responses tailored to individual students, and teacher-in-the-loop refinement of instructional strategies informed by data logged in Jill Watson.

## Adaptation to Course-Specific Content

The first tier of personalization is built on utilizing RAG architecture and vector embeddings of course specific content to develop specialized course knowledge base in Jill Watson compared to general purpose LLMs. This study proposes a human-AI hybrid assessment model to evaluate the effectiveness of Jill Watson in comparison to ChatGPT, leveraging the bidirectional nature of communication between human users and AI agents. A curated set of 20 questions from the Georgia Institute of Technology's CS 7637 Knowledge-Based Artificial Intelligence course will be selected from the Jill Watson repository, which records every question asked by students. Five human experts will evaluate and compared Jill Watson's responses with ChatGPT's responses across several key metrics: accuracy and relevance (Denny et al. 2023; Sari and Alfansi 2024), clarity and readability, engagement and personalization, trustworthiness and safety, and overall helpfulness (Denny et al. 2023).

Currently, the user study has been initiated, and data collection is in progress. Statistical tests such as paired t-tests (Laskar et al. 2024) and ANOVA (Kahng et al. 2024) will be used to test the null hypothesis that Jill Watson is less effective than ChatGPT. Additionally, qualitative analysis such as A/B testing (Quin et al. 2024) will be conducted to assess subjective factors influencing expert preferences. By integrating bidirectional feedback, this research contributes to the evolving landscape of AI-powered educational tools and their role in enhancing student learning outcomes.

## AI-Driven Customization

Customizing Jill Watson's interaction with individual students form the basis of the second tier of personalization proposed in this paper and is focused on the conversational aspect mentioned in the literature (Chen et al. 2024). The present work envisions a real time conversational agent that responds uniquely to individual students utilizing some of the prompt engineering strategies outlined in the paper (Park et al. 2024).

Key advancements in this version of Jill Watson involve serving student assessment surveys and live feedback on personalization collected using the agent interface. Assessment strategies include serving students with baseline and personalized responses and recording their feedback on a Likert scale related to preferred response, accuracy, comprehensiveness, and ease of understanding.

## Teacher-in-the-Loop Refinement

The final tier of personalization proposes a data analytics and visualization pipeline that analyzes Jill Watson's rich data repository to provide insights and personalized recommendations including curated group activities, communication plans, and high-impact examples to teachers

to enable an interactive and supportive learning environment. The Jill Watson dashboard serves as a bridge between raw student interaction data and actionable insights for instructors. By aggregating individualized student data at the micro-level, the dashboard generates class-level insights that evolve into meso-learning patterns, helping teachers identify trends in student engagement and comprehension (Buckingham Shum 2012). These insights provide explanations by visually representing AI-generated data in an intuitive format, allowing instructors to make informed decisions rather than relying on AI in isolation and improve their own teaching (Belghachi 2023). Instead of replacing teachers, the Jill Watson dashboard enhances their ability to personalize instruction by revealing key learning patterns and areas requiring intervention. Looking ahead, we propose expanding these capabilities to include macro-level metrics, enabling educators to analyze trends across entire academic terms and multiple institutions for a broader, more systemic impact on personalized teaching strategies.

## Acknowledgements

This research has been supported by grants from the US National Science (#2247790 and #2112532) to the National AI Institute for Adult Learning and Online Education (aialoe.org).

## References

Belghachi, M. 2023. A review on explainable artificial intelligence methods, applications, and challenges. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, 11(4): 1007–1024.

Buckingham Shum, S. 2012. Learning Analytics: Definitions, Processes, and Potential. Policy Brief 2, UNESCO Institute for Information Technologies in Education. UNESCO IITE Policy Brief No. 2.

Chen, J.; Liu, Z.; Huang, X.; Wu, C.; Liu, Q.; Jiang, G.; Pu, Y.; Lei, Y.; Chen, X.; Wang, X.; et al. 2024. When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web*, 27(4): 42.

Denny, P.; Khosravi, H.; Hellas, A.; Leinonen, J.; and Sarsa, S. 2023. Can we trust AI-generated educational content? comparative analysis of human and AI-generated learning resources. *arXiv preprint arXiv:2306.10509*.

Kahng, M.; Tenney, I.; Pushkarna, M.; Liu, M. X.; Wexler, J.; Reif, E.; Kallarackal, K.; Chang, M.; Terry, M.; and Dixon, L. 2024. Llm comparator: Visual analytics for side-by-side evaluation of large language models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–7.

Kakar, S.; Maiti, P.; Taneja, K.; Nandula, A.; Nguyen, G.; Zhao, A.; Nandan, V.; and Goel, A. 2024. Jill Watson: scaling and deploying an AI conversational agent in online classrooms. In *International Conference on Intelligent Tutoring Systems*, 78–90. Springer.

Laskar, T. R. M.; Alqahtani, S.; Saiful Bari, M.; Rahman, M.; Abdullah Matin Khan, M.; Khan, H.; Jahan, I.; Bhuiyan, A.; Tan, C. W.; Parvez, M. R.; et al. 2024. A Systematic Survey and Critical Review on Evaluating Large Language Models: Challenges, Limitations, and Recommendations. *arXiv e-prints*, arXiv–2407.

Maiti, P.; and Goel, A. 2025. Can an AI Partner Empower Learners to Ask Critical Questions? In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25, 314–324. New York, NY, USA: Association for Computing Machinery. ISBN 9798400713064.

Park, M.; Kim, S.; Lee, S.; Kwon, S.; and Kim, K. 2024. Empowering personalized learning through a conversation-based tutoring system with student modeling. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–10.

Quin, F.; Weyns, D.; Galster, M.; and Silva, C. C. 2024. A/B testing: A systematic literature review. *Journal of Systems and Software*, 112011.

Sari, E. N.; and Alfansi, L. 2024. Elevating satisfaction: Unleashing the power of ChatGPT with personalization, relevance, accuracy, convenience, and tech familiarity. *Manajemen dan Bisnis*, 23(1): 93–106.

Wei, Y.; Jiang, Y.-H.; Liu, J.; Qi, C.; Jia, L.; and Jia, R. 2025. The Advancement of Personalized Learning Potentially Accelerated by Generative AI. In *Society for Information Technology & Teacher Education International Conference*, 991–1000. Association for the Advancement of Computing in Education (AACE).