

# Jiachen Lilian Gong

jgong42@asu.edu    linkedin.com/in/jiachengong/    |    liliangong.github.io    |    Tempe, AZ, USA

## EDUCATION

---

<b>CARNEGIE MELLON UNIVERSITY</b>	Pittsburgh, PA, USA
M.S., Education Technology and Applied Learning Sciences. GPA: 3.91	2019-2020
<b>UNIVERSITY COLLEGE LONDON</b>	London, UK
Data Science Summer School.	2018
<b>PEKING UNIVERSITY</b>	Beijing, China
B.A., Sociology. GPA: 3.72	2014-2018

## EXPERIENCE

---

<b>ARIZONA STATE UNIVERSITY</b>	Tempe, AZ, USA
Research Professional	Oct. 2023 - Present
<ul style="list-style-type: none"><li>• Applied advanced statistical modeling across three institutional studies, including a 13,000+ student longitudinal dataset, to identify behavioral and contextual predictors of academic outcomes; findings shaped ASU retention policy and were published at ACM Learning@Scale.</li><li>• Built two production-grade LLM systems using supervised fine-tuning, retrieval-augmented generation, and prompt engineering: a Personally Identifiable Information (PII) redaction pipeline achieving state-of-the-art F1 across 250,000+ student records, and an adaptive GenAI language tutor validated through multi-phase user evaluation including user survey and interviews.</li><li>• Designed and executed a between-subjects mixed-methods experiment to validate privacy-preserving synthetic datasets for education research; evidence-based findings were directly adopted into the data governance framework of a cross-institutional platform now serving 80 universities.</li></ul>	

## EXPLORELEARNING

---

Data Science Engineer	Remote, USA
	Sept. 2020 - Sept. 2023
<ul style="list-style-type: none"><li>• Engineered ETL pipelines to transform raw student-system interaction logs into analytics-ready datasets, forming the data foundation for all learning analytics products across company's educational game suite.</li><li>• Designed and deployed a Bayesian Knowledge Tracing (BKT) model that powered adaptive content sequencing in K-12 math games, producing personalized learning paths and measurable learning gains.</li><li>• Synthesized learning analytics findings into monthly executive briefings for the CEO and learning design lead, translating complex behavioral patterns into product roadmap decisions and iterative improvements to the adaptive learning system.</li></ul>	

## TECHNICAL SKILLS

---

**Statistical Methods:** multilevel modeling, regression analysis, survival analysis, Bayesian knowledge tracing

**AI/ML:** LLM supervised fine-tuning, retrieval-augmented generation, prompt engineering, ensemble methods

**Tools:** Python (pandas, scikit-learn, PyTorch, TensorFlow), R, SQL, STATA, SPSS

**Research Methods:** statistical analysis, machine learning, experimental design, survey design, interview

## PUBLICATIONS

- 
1. **Gong, J.**, Bali, A., Ahmed, I., Banawan, M., Arner, T. & McNamara, D. Re-imagine Knowledge Tracing with Student Agency in a Generative AI Language Tutor. (Accepted) 27th International Conference on AI in Education (AIED '26).
  2. **Gong, J.**, Goldshtein, M., Xu, X., Arner, T., Roscoe, R. D., & McNamara, D. 12 English and culture as factors in college math achievement. In Proceedings of the Twelfth ACM Conference on Learning@ Scale (L@S '25). ACM. <https://doi.org/10.1145/3698205.3733957>
  3. Imundo, M. N., Goldshtein, M., Watanabe, M., **Gong, J.**, Crosby, D. N., Roscoe, R. D., Arner, T., & McNamara, D. S. (2025). Awareness to Action: Student Knowledge of and Responses to an Early Alert System. Applied Sciences, 15(11), 6316. <https://doi.org/10.3390/app15116316>

4. Imundo, M. N., Watanabe, M., Potter, A. H., **Gong, J.**, Arner, T., & McNamara, D. S. (2024). Expert thinking with generative chatbots. *Journal of Applied Research in Memory and Cognition*, 13(4), 465-484. <https://doi.org/10.1037/mac000199>
5. Christhilf, K., **Gong, J.**, & McNamara, D.S., (2024). Context-embedded knowledge tracing and latent concept detection in a reading game. In D. Spikol, O. Viberg, A. Martinez-Mones, & P. Guo (Eds.), *L@S '24: Proceedings of the eleventh ACM conference on learning @ scale*. Association for Computing Machinery. <https://doi.org/10.1145/3657604.3664674>

## RESEARCH AND DEVELOPMENT PROJECTS

### DEVELOPMENT AND EVALUATION OF LANGUAGE LEARNING LLMs

Directed the development and evaluation of an AI-powered language tutor bot, overseeing system architecture, customized LLM design, and iterative model optimization. Designed and executed a multi-phase quantitative evaluation framework to assess agents' adaptivity, accuracy, and instructional quality; iterated on model design based on quantitative performance signals.

### SUPERVISED FINE-TUNING LLM FOR TEXT REDACTION

Led the end-to-end development of a robust Personally Identifiable Information (PII) removal pipeline: designed the annotation schema, constructed a ground-truth training dataset, fine-tuned an LLM achieving state-of-the-art F1, and deployed the system to redact 250,000+ student discussion board entries, releasing a privacy-safe dataset for the broader education research community.

### A COMPREHENSIVE EVALUATION OF AN EARLY ALERT SYSTEM

Analyzed 13,000+ student institutional records to model the predictive relationship between early alert interventions and academic outcomes; applied multi-level mixed effect regression analysis to quantify that negative alerts are strong predictors of GPA decline and dropout risk, with findings directly driving university-level retention strategy decisions.

## PRESENTATIONS

<b>ASU DATA CONFERENCE</b>	Tempe, AZ, USA
"Personally Identifiable Information (PII) Removal in Discussion Board Data"	2025.11.03
<b>ASU DATA CONFERENCE</b>	Tempe, AZ, USA
"From Status to Success: Quantifying the relations of Academic Status Reports (ASRs), students' academic outcomes, and retention"	2025.11.05
<b>AECT INTERNATIONAL CONVENTION</b>	Las Vegas, NV, USA
"Generative AI in French Language Learning"	2025.10.24
<b>ACM LEARNING AT SCALE CONFERENCE 2025</b>	Palermo, Italy
"Language Equation: 12 English and Culture as Factors in College Math Achievement"	2025.07.23
<b>ASU RESEARCH TECHNOLOGY OFFICE IGNITE: AI</b>	Tempe, AZ, USA
"Personally Identifiable Information (PII) Removal in Discussion Board Data: LLMs Fine-tuning and ETL Pipeline Building"	2025.07.01
<b>ICICLE LEARNING ENGINEERING CONFERENCE</b>	Tempe, AZ, USA
"Leveraging Learning@Scale Data to Uncover Impacts of Discussion Board Use in Higher Ed"	2024.07.23
<b>ACM LEARNING AT SCALE CONFERENCE 2024</b>	Atlanta, GA, USA
"Context-embedded Knowledge Tracing and Latent Concept Detection in a Reading Game"	2024.07.18