Ji Yong Cho

106 Village Circle, Ithaca, NY 14850

Email : jc3374@cornell.edu

Website : https://jiyongcho.github.io/

I research user engagement in online platforms and technologies, including identifying factors associated with engagement, devising interventions to increase engagement, and measuring the impact of the interventions. I primarily take a quantitative approach by applying data science and machine learning techniques, but I also conduct qualitative studies to gain deeper insights into target users during software development.

Education

Cornell University — Ithaca, NY

Ph.D. in Information Science — *Aug. 2018 – Aug. 2024 (expected)*

University of Pennsylvania (UPenn) — Philadelphia, PA

M.C.I.T in Computer Information and Technology — *2018*

M.S.Ed. in Interdisciplinary Studies in Human Development — *2016*

Ewha Womans University — Seoul, South Korea, 2014

B.B.A in Psychology, Business Administration — *2014*

Working Experience

Research scientist (@Applied Machine Learning Lab) — *LG AI Research, April 2023 - June 2024*

LLMs' Performance Metrics in Expert Knowledge Domains

- Research the consistency of LLM auto-evaluation in long-form writing and question answering (QA)
- Research diverging and converging knowledge in LLMs' QA (a manuscript under review)
- Developed a measure for inspiration in the use of LLMs in research ideation (a manuscript under review)

Human-Centered Evaluation of LLMs

- Identified use cases for LLM-powered tools for knowledge workers
- Administered individual interviews and focus group studies with 130+ researchers and engineers in AI/ML area
- Developed an evaluation taxonomy for LLMs' QA capabilities, driven by theoretical work in Social Science and reflecting perspectives from foundational ML researchers
- Designed and ran end-to-end human evaluations on LLMs for 3 cycles of model development
- Analyzed, visualized, and wrote up end-to-end human evaluations for diverse stakeholders

Data Pipelines for LLM development

- Developed a data collection tool for an LLM; API development and wireframe design as a part of a team
- Administered a series of data collection, working with external research labs
- Manage database and curate training/test data for each module in an LLM
- Share insights with the engineering team for improving models from end-users perspectives

Knowledge Sharing Tool Development

- Design and develop an LLM-powered knowledge-sharing tool for researchers
- Write APIs, set up databases, design wireframes
- Test a Proof of Concept by running prototype studies
- Collaborate with external labs in NLP, human-computer interaction, and industrial design for research projects
- Research useful interaction designs in LLM-powered systems
- Prepare manuscripts (under review) and give a demonstration (e.g., NeurIPS 2023)

Skills

- **Languages**: proficient: Python, R; intermediate: SQL, JavaScript; basic: C/C++, Java, Go
- **Development**: FastAPI, MongoDB, MySQL, Docker, Google Cloud Computing
- **Data Analytics**: Log data mining, Causal inference, Online-based experiments, Sampling, Data imputation, Survey & measure development, Python libraries (scikit-learn, etc.), Qualtrics
- **UX/Intervention Studies**: Quantitative research (regression, ANOVA, etc.), Qualitative research (Interviews, focus group studies, Co-design workshops, Persona development/ User scenarios, Figma, Content analysis)

Data Projects

Online Course Learner Profiles *Cornell University, Jan 2021 - May 2023*

- Perform data mining on users' large behavioral time stamp data (92GB), collected from 150+ Harvard EdX Massive Open Online Courses (MOOCs) for 3+ years.
- Find behavioral patterns using AI/ML techniques and connect them with 1.5 billion user surveys (e.g., study plans).

Help-seeking in College Subreddits *Cornell University, Jan 2020 - Dec 2021*

- Investigated topics users shared in online communities (Reddit).
- Identified factors associated with users' active participation in online communities.
- Applied NLP techniques and fitted predictive modeling with 3-year Reddit archived data (100K+ posts) and U.S. Department of Education data (1K+ colleges)

Student Engagement in Corde.org *Cornell University, Jan 2019 - Dec 2020*

- Examined changes in students' engagement with an online learning platform over time.
- Identified important factors that affect students' continuous use of the platform and their learning outcomes.
- Analyzed a year-long trace data of 83K+ students from 11K+ class sections, combined with three-time-point user surveys, imputing missing data, and fitting hierarchical linear models.

Development Projects

CoachEra (Mobile application) *Cornell University, Jan 2020 - Dec 2022*

- Designed and developed a chatbot-based app to help learners make weekly study plans, utilizing open-sourced code.
- Led a team of 5+ members (graphic/UX/UI designers, front/back-end engineers.
- Released the app in iOS & Android to serve 200+ students simultaneously.

Delivery Ghost (Browser web game) *Cornell University, Jan 2018 - Dec 2020*

- Designed and developed a game that teaches Chinese in an immersive way, with minimal English translation.
- Led a team of 5+ members, developers & graphic designers.
- Designed game mechanics, implemented with Phaser engine (JavaScript); connected with MySQL database; conducted user studies.
- The game was played by 500+ learners, demonstrating a significant improvement in vocabulary acquisition.

Music Play E-Textile Learning Kit (Physical computing kit) *UPenn, 2018*

- Developed educational artifacts to teach programming concepts, using circuits and Arduino in C++.
- Worked with learning scientists, instructors, and target students
- The kit was used for a CS summer bootcamp for 50+ high school students

Other Experiences

- **Teaching**: Taught weekly sessions, coached students for semester-long projects, and provided feedback on assignments and projects for 6 courses (50 - 300 students in class), including introductory programming (Java & Python), web development/app prototyping, and learning analytics. Received Cornell Information Science Outstanding TA Award Year 2018-9 for excellence in teaching user-centered web development and guiding students' projects.
- **Research**: Led 5+ collaborative research projects using various methods, including data mining, regression fitting, and predictive modeling. Assisted 5+ collaborators in different fields (e.g., Marketing, communication, education, psychology, design, natural language processing). Published research papers at CHI, L@S, etc.. (publications).
- **Mentoring**: Mentored 20+ undergraduate students in research projects (2018-2023); Served as a mentor for first-year PhD students in Information Science (2019-2023)

Graduate-Level Technical Courses Taken

- **Computer Science**: Software engineering, Artificial intelligence, Big data analytics, Computer system programming, Operating system design and implementation, Networked system, Implementation of cloud networks
- **Statistics**: Regression and analysis of variance, Causal reference, Linear models with matrices, Learning analytics
- **UX/UI research**: Research design and analysis, Computational methods for information science research, Social research design and methods, Human-computer interaction studio