Jiayu (Beatrice) Ye

Durham, NC ● (984)3406523 ● Beatrice.ye@duke.edu ● LinkedIn ● Personal Website ● U.S. Permanent Resident

SUMMARY

Duke sophomore double-majoring in Computer Science and Sociology, with hands-on experience building AI products across early-stage startups and industry teams. Combines technical fluency with an understanding of human behavior to translate ambiguous problems into thoughtful, user-centered products.

EDUCATION

Duke University ***Aug 2025 - May 2029***
A.B. in Computer Science & Sociology GPA: 4.00/4.00
Activities: Duke Girls into VC | Business-Oriented Women | Duke Players | DCSSA

PROFESSIONAL EXPERIENCE

POIZON **Shanghai, China**
AI Product Manager Intern ***Jun 2026 - Aug 2026***

- Owned the 0-1 product design for Gift Agent (a sub-agent within POIZON's Shopping Agent) to support context-aware gift decision-making: defined intent taxonomy, decision factors, dialogue strategies, and retrieval logic, reaching **15K daily UV**
- Translated Gift Agent's product logic into UI design, covering entry-point placement and copy, recommendation cards, and follow-up modules that operationalized core product value and decision logic
- Developed Gift Agent's evaluation pipeline, defining 7 eval metrics and 200-case golden sets to drive iteration; diagnosed issues from online evaluation results, LLM execution traces, and 50+ pre-release user feedback; refined prompting and multi-turn context management to **improve Gift Agent performance by 12.4%**
- Launched Gift Agent, lifting Shopping Agent's **conversations per user by 4.60%** and **products browsed per user by 74.42%**
- Led quality improvement for query decomposition and web search nodes in Shopping Agent's LangGraph pipeline: defined 12 node-level metrics and real-case eval sets via SQL, diagnosed failure patterns from execution traces, and partnered with engineering to prioritize fixes; **improved entry-to-order conversion by 2.1%** and **ordering UV by 1.96%**

Franklink **Remote, United States**
AI Product Manager & Growth Intern ***Nov 2025 – Jan 2026***

- Diagnosed low retention post-MVP launch: conducted market & user research through competitive benchmarking and 20+ semi-open interviews, and identified limited persistent network content and an overly formal Agent experience for Gen Z users
- Redesigned Franklink's retention loop, owned the feature design of a connection-update recommendation system based on profile-fit and interaction-closeness scores, and an adaptive Agent persona; authored PRDs and partnered with engineering to finalize matching algorithm & lead prompt engineering
- Validated the redesign through a before-after usability testing with 30 seed users; **83%** (25/30) preference for the new persona, and **57%** (17/30) cited the connection-recommendation feature as increasing return value; supported a **$50K pre-seed raise**

PROJECT

LLM Decision Drift Evaluation Benchmark **North Carolina, United States**
Society Centered AI Hackathon 2026 (*devpost.com/software/decision-drift-under-risk-contexts-benchmark-for-llms*) ***Jan 2026***

- Identified limitations in LLM safety evaluations for multi-turn interactions; defined decision drift as a context-driven failure mode where models gradually deviate from safety boundaries as user intent evolves
- Developed the conceptual framework for a multi-turn safety benchmark, structuring 5 risk scenario families, corresponding interaction skeleton, and risk trajectory to diagnose alignment loss
- Collaborated on translating the framework into an evaluation methodology for assessing reliability in high-risk AI applications

RESEARCH EXPERIENCE

MADLAB, Kenan Institute for Ethics, Duke University · OpenAI Funded **North Carolina, United States**
Undergraduate Research Assistant ***Mar 2026 - Present***

- Contributing to the Patient Preference Prediction project, developing an information extraction pipeline that transforms unstructured end-of-life care conversations into machine-readable preference representations for downstream modeling
- Designed a schema-guided LLM extraction framework by defining taxonomies across domain features & usages, latent values, decision reasoning, and counterfactual preferences to enable systematic representation of decision-making patterns
- Developed a Python-based prompt engineering pipeline, incorporating structured output, uncertainty-aware confidence scoring, and text-span grounding to improve extraction reliability

LANGUAGE & SKILLS

Skills: Python, Java, R, SQL, Excel, Figma, Canva, Vercel, Squarespace, PowerPoint
Industry Knowledge: PRD Writing, User Research, Requirement Analysis, A/B Testing, Data visualization, Cross-functional Collaboration, AI Evaluation, Prompt Engineering