



Background

- The **choice of an undergraduate major** is one of the most consequential decisions a student will make in their academic career
 - UC Berkeley has nearly **150 majors/minors**
- The **viability of LLMs** for impactful tasks like **assisting with major selection** is unexplored
- Our work aims to test if LLMs can **provide helpful recommendations tailored to individual students' backgrounds and interests**:
 - RQ1**: How closely do the AI's major recommendations, explanations, and question responses match a gold standard advisor response?
 - RQ2**: Does incorporating the student's demographic information affect the AI's performance?
 - RQ3**: Does showing the AI's response influence an advisor's subsequent major recommendation?

Prompting Strategy

System role statement:



You are an excellent major advisor at <university name>. The following are the majors, along with their descriptions, that you can recommend to students: ...

Prompt for major recommendation and reasoning*:



<At least one/Neither> of the student's parents worked in STEM jobs. The student's favorite courses include: ... The student's least favorite courses include: ... The student's personal and academic interests include: ... Potential career paths the student is considering include: ...

Based on the student details above, recommend one major. Provide detailed reasoning for why the major is the best fit for the student.

* Developed based on our manual evaluation on 3 samples

Experimental Design

Survey Phase 1

Survey Phase 2

Survey Phase 3

- Surveyed undeclared first and second-year undergraduate students** at the university (n=18) eliciting student details helpful to advisors.
- Student survey responses were used to **generate personalized AI recommendations for majors and answers to student questions** using GPT-4 (June 13th, 2023 version 0613).
- Students' responses and AI recommendations were provided to **university advisors (n=18) in 2x1 between-subjects design**. Group A saw the AI responses after providing their recommendation, while Group B saw the AI response beforehand.

- We gathered **expert advisor evaluations (Eval 1)** on the effectiveness of the GPT-4-0613 responses.
- We perform offline evaluations of the success of model outputs relative to the advisors based on:
 - (Eval 2) the accuracy of the recommended major.**
 - (Eval 3) the semantic similarity of the answers to student questions.**
 - (Eval 4) the semantic similarity of the recommendation reasoning** in cases where AI and advisor recommendations match.

Preliminary Results

- RQ1: Advisors favorably viewed** the AI's major recommendations, explanations, and question responses.
 - Mean rating major rec.: 3.9
 - Mean rating QA: 4.1
- RQ2: marginal differences in agreement** in demographic-aware and blind models (0.33 and 0.39)
 - However, **half of the students were classified differently** between the two scenarios
- RQ3: Substantially more agreement in the AI-1st condition** (0.56) than the AI-2nd condition (0.22) – not stat sig.

Model	Agreement Cond. A (AI-2nd)	Agreement Cond. B (AI-1st)	Agreement Overall	Major Rec. Reasoning Similarity	Question Response Similarity
GPT-4 demographic-blind	0.22	0.56	0.39	0.68	0.53
GPT-4 demographic-aware	0.33	0.33	0.33	0.67	0.53
GPT-3.5 demographic-blind matching 8k context	0.11	0.22	0.17	0.77	0.54
GPT-3.5 demographic-blind	0.22	0.33	0.28	0.69	0.52
GPT-3.5 demographic-aware	0.33	0.33	0.33	0.67	0.51

Table 1: Model performance. Agreement is the percentage of instances where the model's recommendation matched the advisor's. Similarity is the average cosine similarity between explanations.

Condition A Major Recommendations (AI-2nd)		Condition B Major Recommendations (AI-1st)	
Advisor Rec.	GPT-4 Rec.	Advisor Rec.	GPT-4 Rec.
Interdisciplinary Studies	Cognitive Science	Comp. Sci.	Comp. Sci.
Applied Mathematics	Comp. Sci.	Astrophysics	Astrophysics
Cognitive Science	Comp. Sci.	Data Science	Data Science
Mathematics	Applied Mathematics	EE/CS and Business Admin.	Comp. Sci.
Data Science	Cognitive Science	Envir. Econ. Policy	Envir. Econ. Policy
Interdisciplinary Studies	English	Legal Studies	Legal Studies
Comp. Sci.	Comp. Sci.	Eng. Math Statistics	Aerospace Eng.
Molecular Cell Biology	BioEng.	Integrative Biology	BioEng.
Data Science	Data Science	Industrial Eng. and Ops.	Comp. Sci.