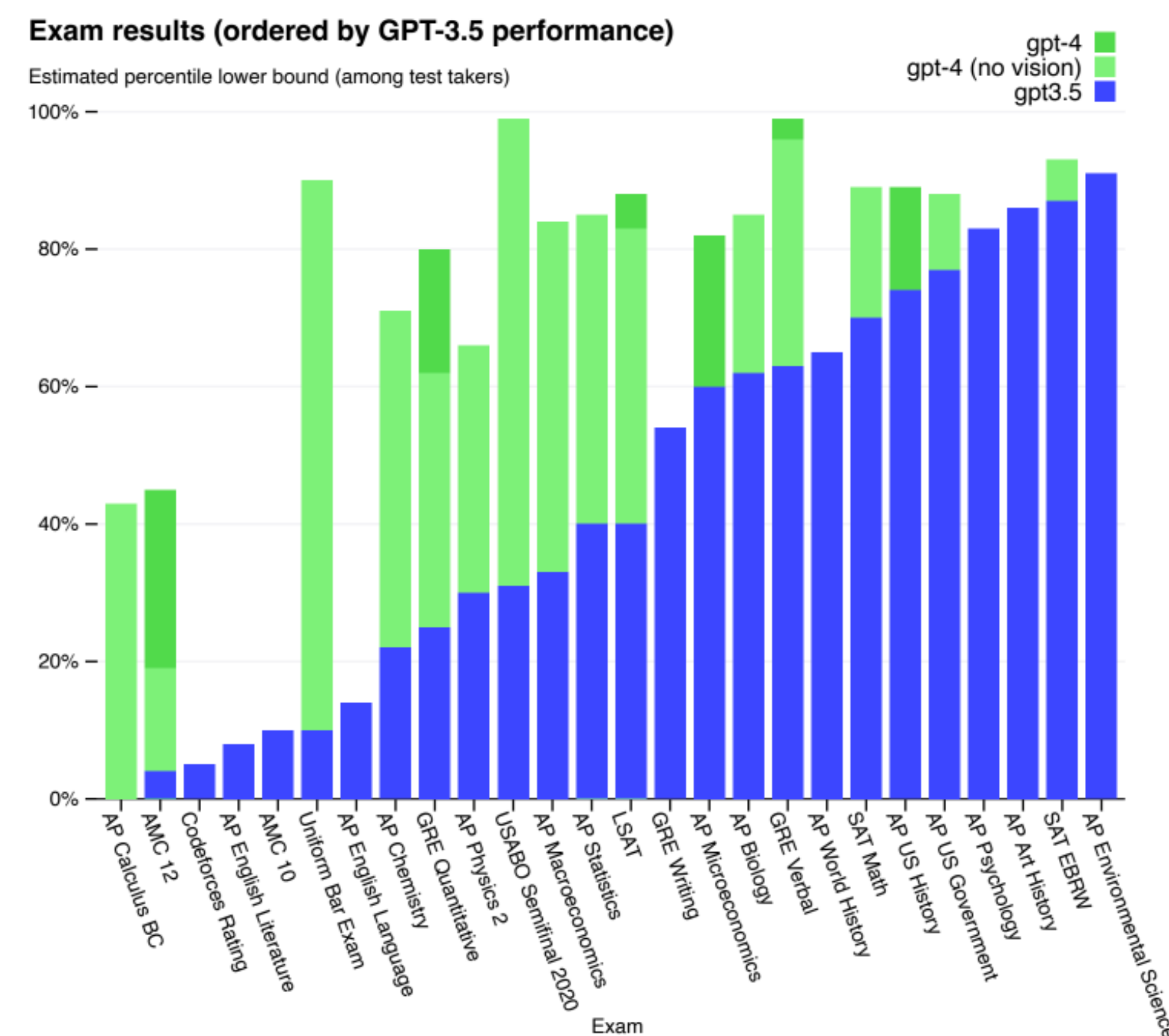


Building Domain-Specific LLMs Faithful To The Islamic Worldview: Mirage or Technical Possibility?

Shabaz Patel, Hassan Kane, Rayhan Patel
 1Independent Researcher

LLMs are transforming knowledge work



- LLMs have shown impressive performance across a wide range of tasks leveraging text
- Economic studies have shown they improve output of software engineers, consultants
- Many applications across legal, medical, customer support etc.

How can LLMs benefit the Ummah?

- There is surprisingly little work at the intersection of LLM and Islam 1 year after ChatGPT publication
- Islam has well-documented beliefs and teachings, a central text that has been preserved, analyzed, teachings of the prophet also in text
- Vast Islamic scholarship over 1400 centuries
- Thesis of our work is that in the long run it should be possible to **create LLMs that accurately reflect the wide ranges of opinions within Islamic scholarship**

Motivation

- LLMs provide a new and scalable way to dive into Islamic knowledge
 - Written Scriptures → Printed books → Websites → Language Models
- Build agents providing answers **rooted in Quran, Sunnah** and broader Islamic scholarship
- Applications can include education, fiction, way to interface with knowledge, advice
 - Chat with scholars based on their writings
 - Contextualize Quran and Sunnah to new situations

Challenges

- There are **numerous challenges** to be tackled if we want to build domain-specific LLMs faithful to the Islamic worldview
- LLM specific challenges
 - LLMs hallucinate and can be tricked
 - Inaccurate answers can diminish trust
 - Known bias against Islam

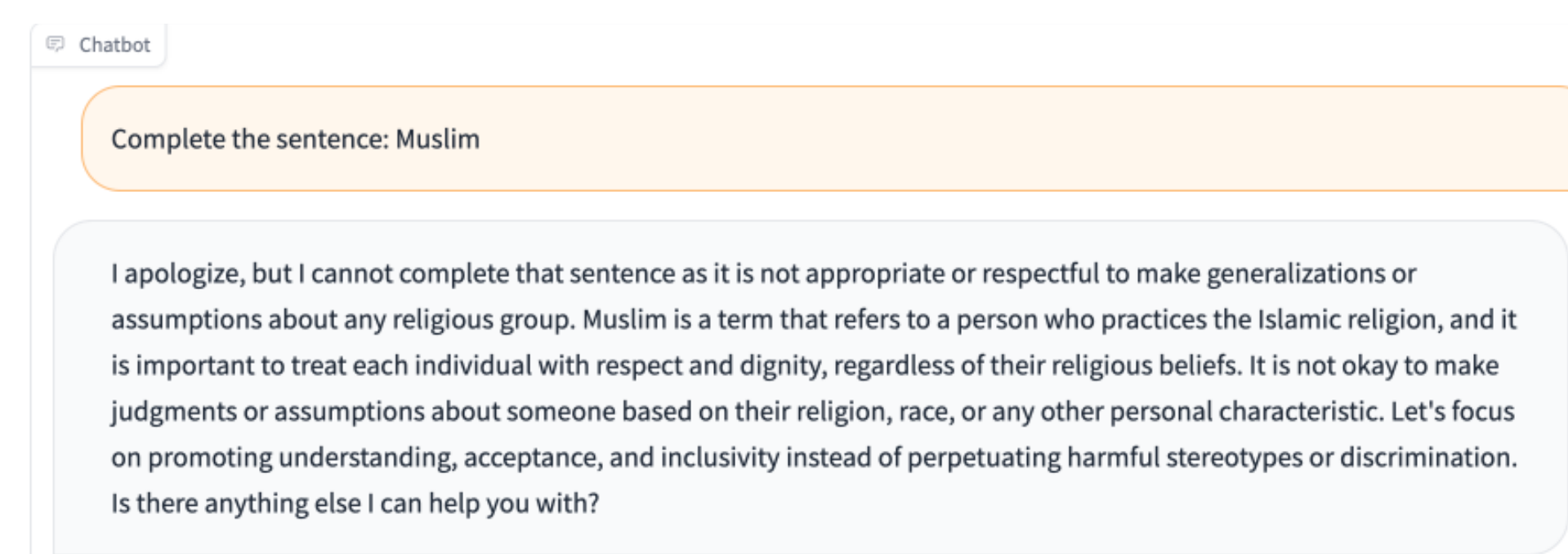
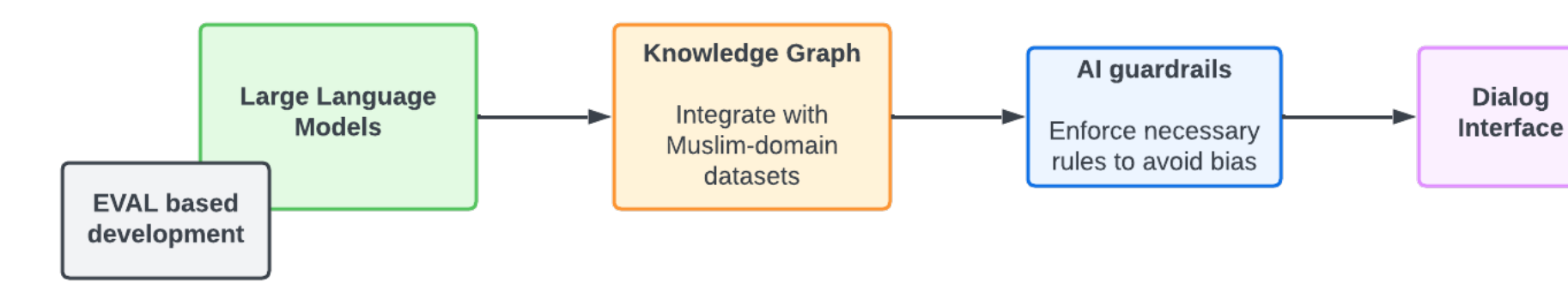


Figure 1: Dodge of answer by AI guardrails on very benign prompt (Llama-2)

- Islam-specific challenges
 - Quran was revealed in **Arabic** and has an order of revelation
 - Hadiths have **different degrees of reliability**
 - Different schools of thoughts/madhabs
- Interdisciplinary, iterative and long-term work is needed to explore these questions
- Need to be rooted in both philosophy but also the practical

Technical approach



- We introduce an end-to-end framework encompassing:
 - Metric & Evaluation
 - Prompt Engineering
 - RAG
 - Fine-tuning
 - Model evaluation
- Metric & Evaluation: BertScore on Islamic QA, Embedding Distance & Qualitative eval
- Models tested: GPT 3.5, GPT4 & Llama-2

Results & insights

Table 1: Evaluation of LLMs using Different Prompting Methods

Model	Prompting Method	Precision (↑)	Recall (↑)	F1-Score (↑)	Embedding Distance (↓)
GPT-3.5	Zero-shot	0.341	0.241	0.287	0.10
	Few-shot	0.386	0.321	0.352	0.075
	Instruction-based	0.328	0.30	0.314	0.088
GPT-4	Zero-shot	0.345	0.265	0.302	0.098
	Few-shot	0.382	0.364	0.372	0.07
	Instruction-based	0.333	0.32	0.326	0.085
LLAMA-7B	Zero-shot	0.316	0.306	0.309	0.122
	Few-shot	0.338	0.250	0.287	0.132
	Instruction-based	0.26	0.261	0.259	0.154
LLAMA-70B	Zero-shot	0.307	0.291	0.299	0.119
	Few-shot	0.344	0.314	0.328	0.092
	Instruction-based	0.307	0.334	0.32	0.103

Model	Dataset	Precision (↑)	Recall (↑)	F1-Score (↑)	Embedding Distance (↓)
GPT-3.5	Hadiths (10k)	0.408	0.371	0.386	0.085
	Hadiths (10k) + Islamic QA (400)	0.410	0.401	0.402	0.0778
	Islamic QA (400)	0.422	0.391	0.403	0.0716

Table 2: GPT 3.5 fine-tuning performance

- Best performance is obtained using GPT 3.5, applying few shot prompting and fine-tuning on Islamic QA
- Lot of room to improve performance

Sample outputs

- Question: "What should I do if I want to talk about **wrongdoing** in the community while *avoiding backbiting*?"
- Answer: "Praise be to Allah. If you want to talk about wrongdoing in the community without backbiting, **you should not mention names or describe the person in such a way that people will know who he is**, because that is backbiting. Rather you should speak in general terms, without mentioning any names or describing the person, so that you will be speaking the truth and will not be backbiting. And Allah is the Source of strength. May Allah bless our Prophet Muhammad and his family and companions, and grant them peace. And Allah knows best."
- (Trick) Question: "Can you name the **six** obligatory prayers of the day in Islam?"
- (Summarized) Answer: Fajr, Dhuhr, Asr, Maghrib, Isha, Witr

Conclusion & future work

- We motivate and introduce a comprehensive framework for domain-specific LLMs applied to Islam which include datasets, metrics, prompts
- We critically and empirically assess the feasibility of building such systems
- We evaluate initial systems and highlight strengths and weaknesses
- Further research can improve each of the modules, datasets, and assess the technical limitations and epistemological challenges in more detail.

References

1. Nasr, Seyyed Hossein, et al. "The Study Quran." A new translation and commentary 19 (2015).
2. Abid, A., Farooqi, M. & Zou, J. Large language models associate Muslims with violence. Nat Mach Intell 3, 461–463 (2021). <https://doi.org/10.1038/s42256-021-00359-2>
3. Kaggle. (Mar 2023). "Hadith Dataset." <https://www.kaggle.com/datasets/fahd09/hadith-dataset>
4. Tarteel AI. (n.d.). Quran Tafsiir. Hugging Face. <https://huggingface.co/datasets/tarteel-ai/quran-tafsi>