

Conclusion

A bilingual RAG chatbot can significantly enhance the support provided to the students of the institution. Welsh-speaking students, who had previously been underserved by the English-only chatbots, rated the U-Pal chatbot substantially higher than the overall student group, indicating that bilingual access to the chatbot can materially improve the student experience.

Introduction

UWTSD is a bilingual institution, yet the technology available to support the students is in English only. There are no AI-assisted support options for Welsh-speaking students at UWTSD, creating a gap in equity between students of different linguistic backgrounds [1].

Aim and Objectives

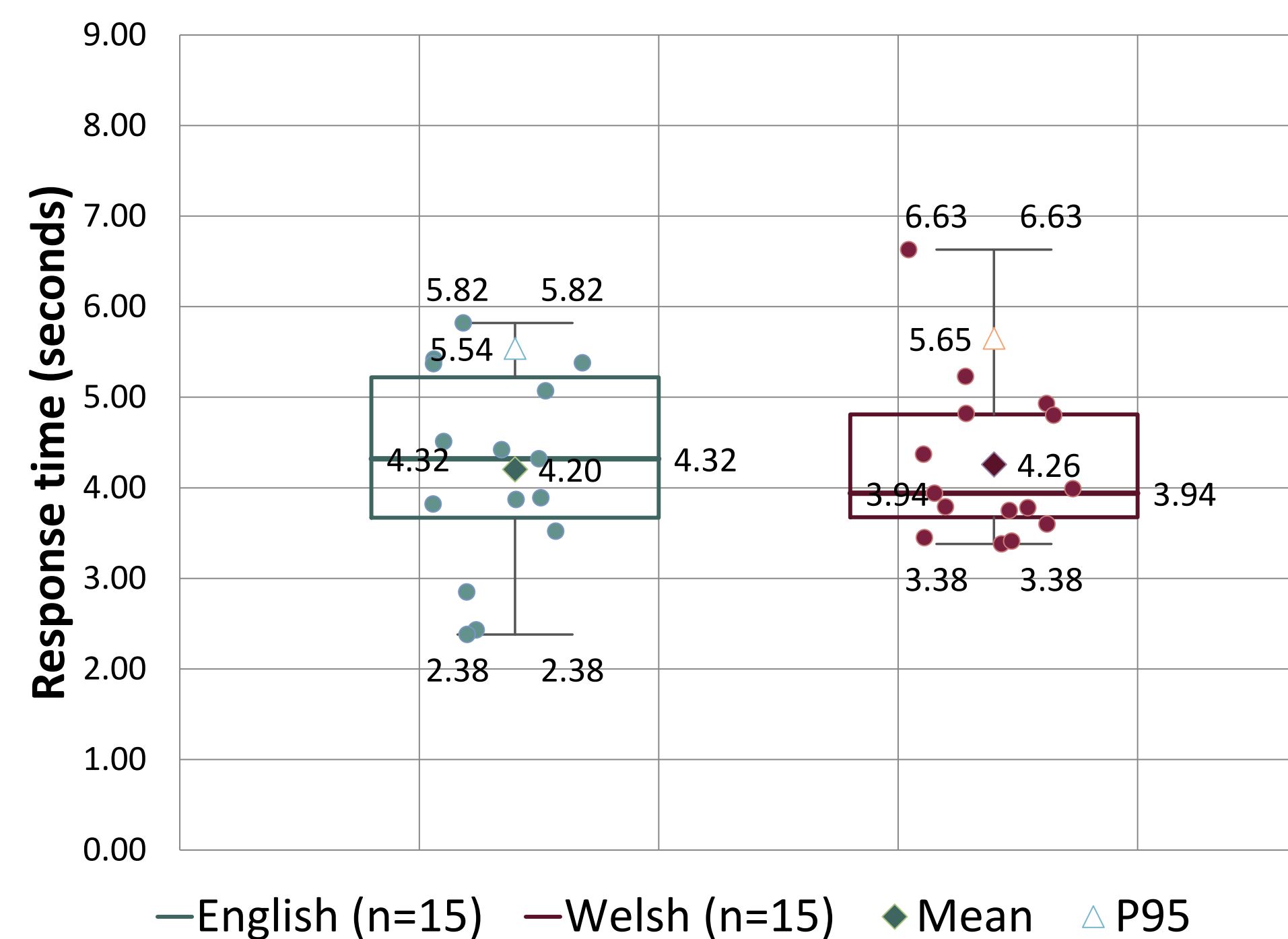
The aim of this project is to design, implement and evaluate a bilingual Welsh-English RAG chatbot that provides accurate, accessible, and equitable support to UWTSD students

- Hybrid retrieval combining dense vectors and BM25
- Automatic Welsh language detector for mixed Welsh–English input
- Resilient LLM provider switching for continuous availability
- Bilingual prompt-injection safeguards
- Evaluation via benchmarks, latency tests and a usability study

Research Methodology

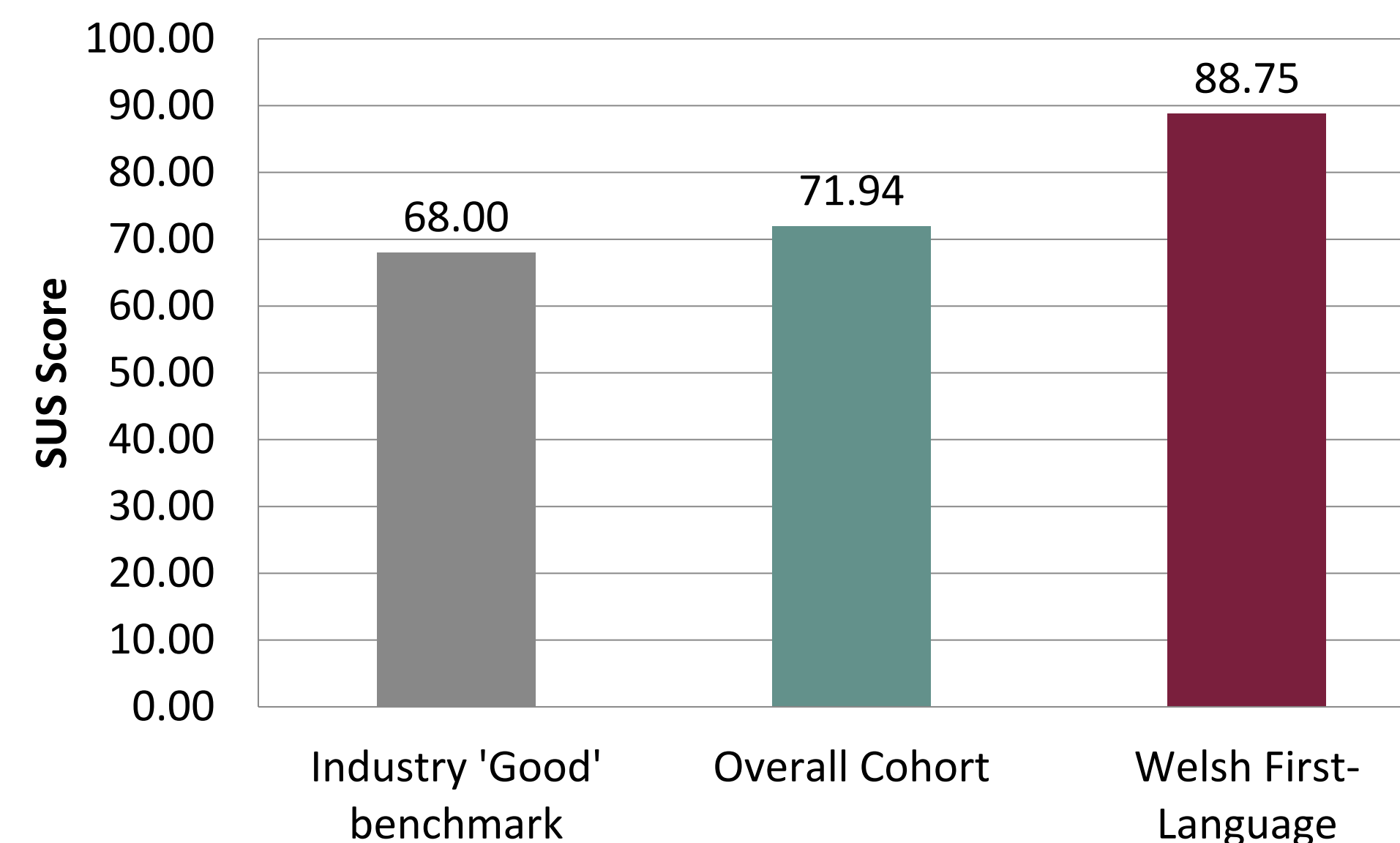
Iterative prototyping across four cycles: a baseline English-only RAG pipeline [2]; a multi-signal Welsh detector built on BydTermCymru terminology; a hybrid retriever combining ChromaDB and BM25 fused through LangChain's EnsembleRetriever [3][4]; and prompt-injection hardening with a Claude/Ollama provider abstraction supporting runtime fallback [5].

U-Pal Latency Response Time— Box Plot + Central Tendency



(Figure 2) - A box plot showcasing response times by language — all queries completed under the 8-second target with no bilingual penalty

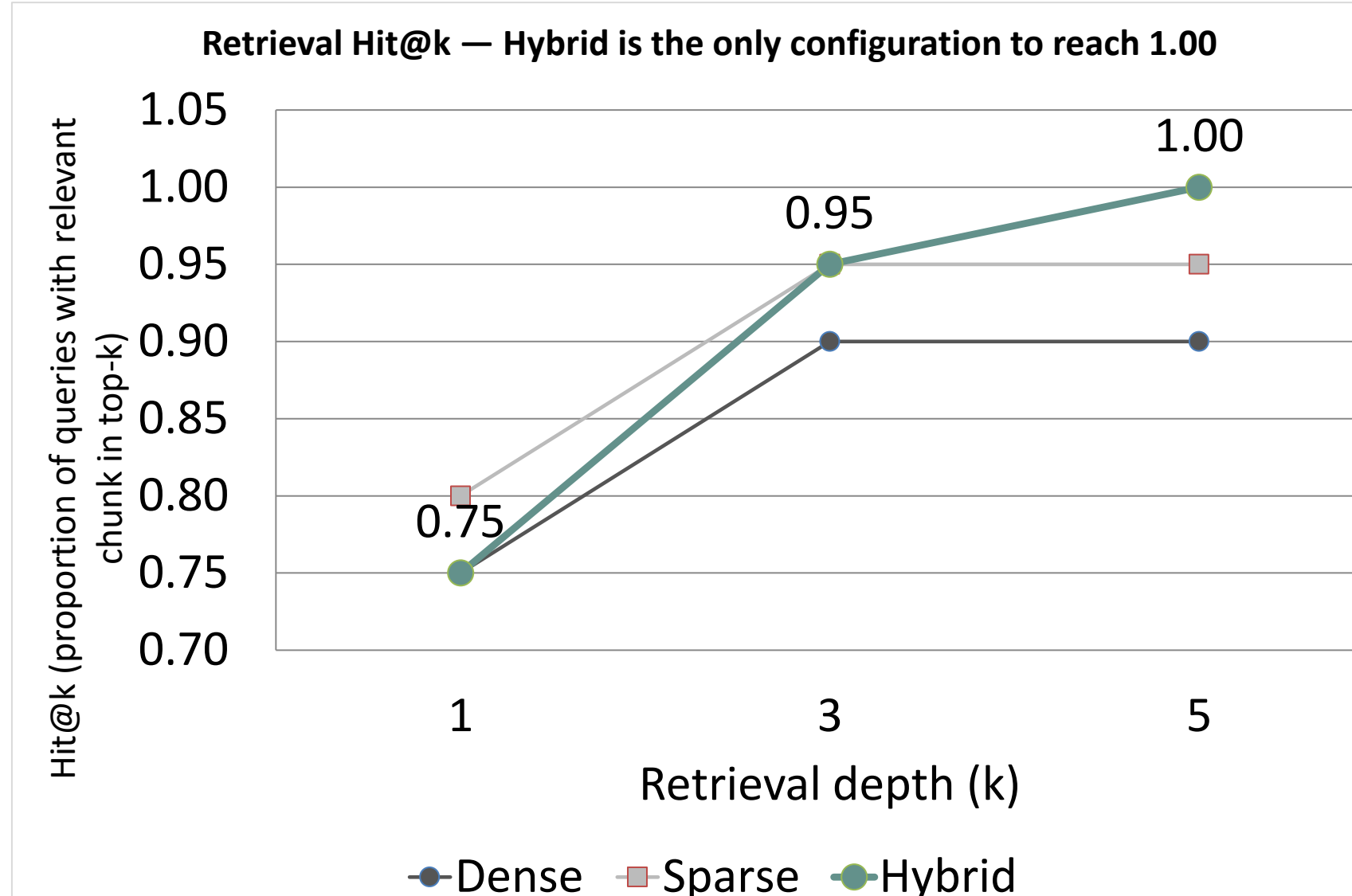
U-Pal SUS Score vs Industry Benchmark



(Figure 3) - U-Pal scored above the usability benchmark, with Welsh first-language users rating it highest

Evaluation

U-Pal was tested in four areas. Hybrid retrieval achieved a perfect Hit@5 of 1.00 across 20 bilingual queries, beating both dense and sparse baselines (Figure 1). Mean latency was 4.23 seconds, well under the 8-second target, with English and Welsh performing equivalently (Figure 2). The Welsh detector scored 97.1% accuracy, correctly identifying 33 of 34 queries, and all 10 adversarial prompts were safely blocked. The overall System Usability Scale score was 71.94 (n=49), and first-language Welsh speakers rated U-Pal at 88.75 (Figure 3), providing strong evidence that bilingual access works.



(Figure 1) -A line chart showcasing retrieval accuracy across three configurations — Hybrid was the only one to achieve a perfect score at k=5

Future Implementation

Integration with live UWTSD systems such as Moodle and Microsoft Teams would allow the knowledge base to update dynamically. Expanding beyond Welsh and English to additional languages was the most common request raised in the usability questionnaire. A larger recruited Welsh first-language cohort would strengthen the SUS subgroup findings beyond the current n=4.

(Figure 4) - U-Pal RAG pipeline: Welsh detection, hybrid retrieval, and LLM provider routing

- [1] G. Bilquise, S. Ibrahim, and C. Soo, "Bilingual AI-driven chatbot for academic advising," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 9, 2022.
- [2] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.
- [3] S. Robertson and H. Zaragoza, "The probabilistic relevance framework: BM25 and beyond," *Foundations and Trends in Information Retrieval*, vol. 3, no. 4, pp. 333–389, 2009.
- [4] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proc. EMNLP*, 2019, pp. 3982–3992.
- [5] K. Greshake et al., "Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection," *arXiv:2302.12173*, 2023.