BÁO CÁO BÀI TẬP: ĐÁNH GIÁ TÀI LIỆU HỌC THUẬT

Chủ đề: Ứng dụng của Mô hình Ngôn ngữ Lớn (Large Language Models - LLMs) trong Kỹ nghệ Phần mềm (Software Engineering)

I. GIỚI THIỆU CHỦ ĐỀ

Trong bối cảnh ngành Khoa học máy tính phát triển mạnh mẽ, các Mô hình Ngôn ngữ Lớn (LLMs) như GPT-4, Llama, và các công cụ phái sinh như GitHub Copilot đang tạo ra một cuộc cách mạng trong Kỹ nghệ Phần mềm. Chủ đề **"Ứng dụng của Mô hình Ngôn ngữ Lớn (LLMs) trong Kỹ nghệ Phần mềm"** tập trung vào việc nghiên cứu cách các AI này hỗ trợ lập trình viên trong các khâu: sinh mã tự động (code generation), phát hiện lỗi (bug detection), viết tài liệu (documentation) và kiểm thử (testing).

Đây là một chủ đề mang tính thời sự cao, đòi hỏi phải tham khảo các nguồn tài liệu cập nhật nhất từ các tạp chí khoa học chuyên ngành, hội nghị đầu ngành (CHI, ICSE, ACM) và báo cáo kỹ thuật từ các tập đoàn công nghệ.

II. QUÁ TRÌNH TÌM KIẾM THÔNG TIN

Quá trình tìm kiếm tài liệu được thực hiện thông qua nhiều kênh khác nhau để đảm bảo tính đa chiều và độ tin cậy:

1. **Cơ sở dữ liệu học thuật:** Sử dụng *Google Scholar*, *IEEE Xplore* và *ACM Digital Library* với các từ khóa *"Large Language Models in Software Engineering"*, *"AI pair programming"*, *"Code generation LLMs evaluation"*.
2. **Tạp chí khoa học & Kỷ yếu hội nghị:** Tìm kiếm các bài báo đã qua phản biện (peer-reviewed) tại các hội nghị hàng đầu như *CHI Conference*, *NeurIPS*, và tạp chí *IEEE Software*.
3. **Sách chuyên khảo:** Tìm kiếm các giáo trình chuẩn mực về Xử lý ngôn ngữ tự nhiên (NLP) để làm nền tảng lý thuyết.
4. **Các nguồn mở trên Internet:** Tra cứu báo cáo kỹ thuật từ các tổ chức phát triển AI (OpenAI) và khảo sát thực tế từ cộng đồng lập trình viên (Stack Overflow).

III. ĐÁNH GIÁ ĐỘ TIN CẬY CỦA TÀI LIỆU

Các tài liệu được đánh giá dựa trên 5 tiêu chí:

- **Tác giả:** Trình độ học vấn, chuyên môn và cơ quan công tác (Ví dụ: Đại học Stanford, Microsoft Research, OpenAI).
- **Cơ quan xuất bản:** Mức độ uy tín của nhà xuất bản (IEEE, ACM, Springer) hoặc nền tảng lưu trữ (arXiv, tuy nhiên cần xét kỹ vì arXiv chưa qua peer-review).
- **Phương pháp nghiên cứu:** Tính chặt chẽ, có sử dụng số liệu thực nghiệm (empirical studies) hay khảo sát hệ thống (systematic review) rõ ràng không.

- **Trích dẫn:** Số lượng trích dẫn cho thấy tầm ảnh hưởng của tài liệu trong cộng đồng học thuật.
- **Tính cập nhật:** Vì AI là lĩnh vực phát triển nhanh, các tài liệu từ năm 2022-2024 được ưu tiên và đánh giá cao nhất.

IV. BẢNG TỔNG HỢP VÀ XẾP HẠNG ĐỘ TIN CẬY CỦA TÀI LIỆU

STT	Tên tài liệu / Tác giả (Năm)	Loại tài liệu	Đánh giá theo các tiêu chí (Tác giả, NXB, PP Nghiên cứu, Trích dẫn, Cập nhật)	Xếp hạng độ tin cậy
1	Large language models for software engineering: A systematic literature review / Hou et al. (2023)	Bài báo khoa học (Tạp chí ACM)	**Tác giả:** Nhóm nghiên cứu từ ĐH Monash. **NXB:** ACM - rất uy tín. **PP Nghiên cứu:** Tổng quan tài liệu hệ thống (SLR) rất chặt chẽ. **Trích dẫn:** Cao (>200). **Cập nhật:** Rất mới (2023).	**Rất cao**
2	Expectation vs. experience: Evaluating usability of code generation tools / Vaithilingam et al. (2022)	Bài báo Khoa học (Kỷ yếu hội nghị CHI)	**Tác giả:** Chuyên gia từ Harvard và ĐH Washington. **NXB:** ACM (Hội nghị CHI hàng đầu về HCI). **PP Nghiên cứu:** Thử nghiệm người dùng (User study) định lượng và định tính. **Trích dẫn:** Khá cao. **Cập nhật:** Tốt (2022).	**Rất cao**
3	Training language models to follow instructions with human feedback / Ouyang et al. (2022)	Bài báo Khoa học (Kỷ yếu NeurIPS)	**Tác giả:** Các nhà khoa học từ OpenAI. **NXB:** NeurIPS (Hội nghị AI số 1 thế giới). **PP Nghiên cứu:** Thực nghiệm thuật toán (RLHF). **Trích dẫn:** Rất lớn (>8000). **Cập nhật:** Tốt (2022).	**Rất cao**

4	Prompting for software engineering / Treude (2023)	Tạp chí chuyên ngành	**Tác giả:** TS. Christoph Treude (ĐH Melbourne). **NXB:** IEEE Software. **PP Nghiên cứu:** Phân tích lý thuyết và thực hành. **Trích dẫn:** Đang tăng nhanh. **Cập nhật:** Rất mới (2023).	**Cao**
5	Sparks of artificial general intelligence: Early experiments with GPT-4 / Bubeck et al. (2023)	Bài báo Khoa học (Preprint)	**Tác giả:** Microsoft Research. **NXB:** arXiv (Preprint, chưa peer-review chính thức nhưng rất nổi tiếng). **PP Nghiên cứu:** Thực nghiệm phân tích định tính. **Trích dẫn:** Rất cao. **Cập nhật:** Mới (2023).	**Cao**
6	Experiences from showing AI-generated code explanations to SE students / MacNeil et al. (2022)	Bài báo Khoa học (Hội nghị ACM)	**Tác giả:** Các học giả nghiên cứu giáo dục CS. **NXB:** ACM ITiCSE. **PP Nghiên cứu:** Phân tích thực nghiệm trong môi trường lớp học. **Trích dẫn:** Khá. **Cập nhật:** Tốt.	**Cao**
7	Speech and Language Processing (3rd ed. draft) / Jurafsky & Martin (2023)	Sách chuyên khảo	**Tác giả:** Dan Jurafsky (Stanford) - Chuyên gia hàng đầu NLP. **NXB:** Tự xuất bản bản	**Rất cao**

			nháp chờ in ấn. **PP Nghiên cứu:** Tổng hợp lý thuyết chuẩn mực. **Trích dẫn:** Kinh điển (>10,000). **Cập nhật:** Bản nháp 2023.	
8	Deep Learning for Software Engineering / Allamanis et al. (2018)	Bài báo Khoa học (Tạp chí ACM)	**Tác giả:** Miltiadis Allamanis (Microsoft Research). **NXB:** ACM Computing Surveys. **PP Nghiên cứu:** Khảo sát toàn diện (Survey). **Trích dẫn:** Rất cao (>2000). **Cập nhật:** 2018 (Hơi cũ so với tốc độ ra mắt LLM hiện nay).	**Trung bình - Cao** (Bị trừ điểm tính cập nhật)
9	GPT-4 Technical Report / OpenAI (2023)	Báo cáo kỹ thuật (Nguồn mở)	**Tác giả:** Tập đoàn OpenAI. **NXB:** Nguồn mở của công ty. **PP Nghiên cứu:** Báo cáo thông số kỹ thuật và benchmark. **Trích dẫn:** Rất cao. **Cập nhật:** Mới nhất.	**Cao** (Có thể mang tính thương mại/PR)
10	2023 Developer Survey / Stack Overflow (2023)	Nguồn mở Internet	**Tác giả:** Stack Overflow. **NXB:** Stack Overflow. **PP Nghiên cứu:** Khảo sát thống kê cộng đồng (>90,000 người).	**Trung bình** (Dữ liệu thống kê tốt nhưng không phải nghiên cứu học thuật sâu)

			Trích dẫn: Dùng nhiều trong báo chí. **Cập nhật:** Rất mới.	

V. TỔNG KẾT

Qua quá trình thu thập và đánh giá 10 nguồn tài liệu, có thể thấy lĩnh vực nghiên cứu về LLMs trong Kỹ nghệ phần mềm đang bùng nổ. Đa số các nguồn tài liệu có độ tin cậy **Cao** đến **Rất cao**, đặc biệt là các công bố từ các hội nghị như ACM, IEEE và NeurIPS. Một số tài liệu như Báo cáo kỹ thuật của OpenAI hay Khảo sát của Stack Overflow tuy không phải là bài báo học thuật truyền thống nhưng lại cung cấp số liệu thực tế quan trọng, có giá trị ứng dụng cao cho bài nghiên cứu tổng quan.

DANH MỤC TÀI LIỆU THAM KHẢO (Theo định dạng Harvard)

1. Allamanis, M., Barr, E.T., Devanbu, P. và Sutton, C., 2018. A survey of machine learning for big code and naturalness. *ACM Computing Surveys (CSUR)*, 51(4), pp.1-37.
2. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S. và Nori, H., 2023. Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint arXiv:2303.12712*.
3. Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J. và Wang, H., 2023. Large language models for software engineering: A systematic literature review. *ACM Transactions on Software Engineering and Methodology*, 33(5), pp.1-41.
4. Jurafsky, D. và Martin, J.H., 2023. *Speech and language processing*. 3rd ed. draft. [online] Available at: https://web.stanford.edu/~jurafsky/slp3/.
5. MacNeil, S., Tran, A., Hellina, D., Coblenz, M., Ragan, P., March, C., ... & Klaus, M., 2022. Experiences from showing AI-generated code explanations to software engineering students. Trong: *Proceedings of the 28th ACM Conference on Innovation and Technology in Computer Science Education*. Dublin, Ireland, 8-13 tháng 7 năm 2022. New York: ACM, pp.1-7.
6. OpenAI, 2023. *GPT-4 Technical Report*. [online] Available at: https://arxiv.org/abs/2303.08774
7. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. và Schulman, J., 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, pp.27730-27744.
8. Stack Overflow, 2023. *2023 Developer Survey*. [online] Available at: https://survey.stackoverflow.co/2023/
9. Treude, C., 2023. Prompting for software engineering. *IEEE Software*, 41(1), pp.22-26.

10. Vaithilingam, P., Zhang, T. và Glassman, E.L., 2022. Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models. Trong: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. New Orleans, LA, USA, 29 tháng 4 - 5 tháng 5 năm 2022. New York: ACM, pp.1-7.