



## Overview Research on the Use of Large Language Models (LLMs) in the Field of Software Engineering

---

Việt Nguyễn Văn, Khánh Nguyễn Hữu, Duyên Nguyễn Thị,  
Huế Lương Thị Minh, Vịnh Nguyễn Thế and Công Nguyễn Hữu

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

September 17, 2024

# Nghiên cứu tổng quan về sử dụng các mô hình ngôn ngữ lớn (Large Language Models- LLMs) trong lĩnh vực kỹ thuật phần mềm

## Overview research on the use of Large Language Models (LLMs) in the field of software engineering

Nguyễn Văn Việt<sup>1</sup>, Nguyễn Hữu Khánh<sup>2</sup>, Nguyễn Thị Duyên<sup>1</sup>, Lương Thị Minh Huệ<sup>1</sup>, Nguyễn Thế Vịnh<sup>1</sup>, Nguyễn Hữu Công<sup>2</sup>

<sup>1</sup>Trường Đại học Công nghệ thông tin và truyền thông, <sup>2</sup>Đại học Thái Nguyên  
Email: [nvviet@ictu.edu.vn](mailto:nvviet@ictu.edu.vn)

### Abstract

Large Language Models (Large Language Models - LLMs) are large deep learning models, pre-trained with large amounts of data. This study provides a comprehensive survey of LLM adoption in the field of software engineering from January 2017 to December 2023. We explore the emerging field of Large Language Modeling in Software Engineering and propose extended research formulations on applying LLMs to solve technical problems facing software engineers. must face to face. The outstanding characteristics of LLMs bring creativity and innovation to many applications in Software Engineering scope, including installer, design, requirements management, debugging, regenerative architecture, optimization performance, chemical documentation and analysis. However, these emerging properties also pose significant technical challenges; We need techniques that can reliably eliminate incorrect solutions, given limitations such as simulation bias. This survey highlights the important role of developing and developing a reliable and effective LLMs based Software Engineering exploitation.

### Keywords

Keywords: Large Language Models; Software Engineering; Transformer; AI; PRISMA.

### Tóm tắt

Các mô hình ngôn ngữ lớn (Large Language Models - LLMs) là mô hình học sâu quy mô lớn, được huấn luyện trước trên lượng dữ liệu khổng lồ. Nghiên cứu này cung cấp một khảo sát toàn diện về việc ứng dụng LLMs trong lĩnh vực kỹ thuật phần mềm từ tháng 1/2017 đến tháng 12/2023. Chúng tôi khám phá lĩnh vực mới nổi Mô hình Ngôn ngữ Lớn trong Kỹ thuật Phần mềm (LLM4SE) và đề ra những thách thức nghiên cứu mở rộng khi áp dụng LLMs để giải quyết các vấn đề kỹ thuật mà các kỹ sư phần mềm phải đối mặt. Đặc tính nổi bật của LLMs mang lại tính sáng tạo và đột phá cho nhiều ứng dụng trong phạm vi Kỹ thuật Phần mềm, bao gồm lập trình, thiết kế, quản lý yêu cầu, sửa lỗi, tái cấu trúc, tối ưu hiệu năng, tài liệu hóa và phân tích. Tuy nhiên, những đặc tính mới nổi này cũng đặt ra những thách thức kỹ thuật đáng kể; chúng ta cần các kỹ thuật có thể đáng tin cậy loại bỏ các giải pháp không chính xác, chẳng hạn như hiện tượng mô phỏng sai lệch. Khảo sát này nhấn mạnh vai trò quan trọng của việc phát triển và triển khai Kỹ thuật Phần mềm dựa trên LLMs đáng tin cậy và hiệu quả.

### Chữ viết tắt

LLMs	Large Language Models
SE	Software Engineering
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RNN	Recurrent Neural Network
AI	Artificial Intelligence

### 1. Phần mở đầu

Các mô hình ngôn ngữ lớn (LLMs) có tác động nhiều đến các lĩnh vực về công nghệ thông tin trong đó bao gồm cả kỹ thuật phần mềm[1], [2]. Nhiều nghiên cứu và các xuất bản về mô hình ngôn ngữ lớn trong lĩnh vực kỹ thuật phần mềm với nhiều cách tiếp cận, giải pháp khác nhau. LLMs trong lĩnh vực SE đang làm mờ ranh giới ngôn ngữ giữa con người tạo và ngôn ngữ do máy tạo ra.

Bài báo này nhằm mục đích nghiên cứu tổng quan về việc sử dụng các mô hình ngôn ngữ lớn (LLMs) trong lĩnh vực kỹ thuật phần mềm. Các nhà phát triển phần mềm, kiểm thử phần mềm có thể sử dụng LLMs ứng dụng trong các pha phát triển, xây dựng. Sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm là cơ hội để giúp cho các công ty phát triển phần mềm, các nhóm nghiên cứu triển khai được nhanh chóng, giảm thiểu nguồn nhân lực, tài chính.

Để giải quyết các vấn đề nêu trên, các nhóm nghiên cứu, nhà khoa học đã nêu, trình bày tiềm năng, cơ hội, khó khăn, thách thức đối với việc sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm. Điều này được thể hiện thông qua nhiều bài báo nghiên cứu tổng quan, ứng dụng liên quan đến LLMs trong lĩnh vực kỹ thuật phần mềm như: Nhóm tác giả Lenz Belzner, Thomas Gabor & Martin Wirsing, năm 2023 với nghiên cứu “Kỹ thuật phần mềm hỗ trợ mô hình ngôn ngữ lớn: Triển vọng, thách thức và nghiên cứu điển hình”[1], đánh giá các mô hình ngôn ngữ lớn được đào tạo về Code[2], Khả năng nổi bật của các mô hình ngôn ngữ[6], Hướng tới

sự hiểu biết về các mô hình ngôn ngữ lớn trong các nhiệm vụ công nghệ phần mềm[9].

Do vậy, bài báo này nhằm giải quyết vấn đề trên bằng cách phân tích thư mục các bài báo khoa học về việc sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm:

Câu hỏi 1: Số lượng bài báo nghiên cứu trong cơ sở dữ liệu tạp chí, hội thảo đối với việc sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm thay đổi như thế nào?

Câu hỏi 2: Những từ khóa nào xuất hiện thường xuyên nhất về sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm của kho dữ liệu tạp chí, hội thảo?

Câu hỏi 3: Những chủ đề nghiên cứu quan trọng nhất trong việc sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm là gì?

Câu hỏi 4: Những khoảng trống và lĩnh vực cho nghiên cứu trong tương lai là gì?

Việc trả lời các câu hỏi nghiên cứu ở phía trên sẽ giúp cho các nhà phát triển phần mềm có được những góc nhìn cơ bản để tiếp cận theo LLMs trong lĩnh vực kỹ thuật phần mềm. Các nhà nghiên cứu mới có thể tìm được những định hướng nghiên cứu mới trong tương lai thông qua các khoảng trống trong nghiên cứu.

## 2. Phương pháp nghiên cứu

Bài viết sử dụng phương pháp nghiên cứu tổng quan PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). PRISMA cung cấp một kế hoạch làm việc để thực hiện phân tích các bài báo khoa học một cách có hệ thống và minh bạch, giúp nghiên cứu viên có thể đánh giá độ tin cậy của các nghiên cứu trước khi sử dụng chúng trong nghiên cứu của mình.

### 2.1 Nguồn tìm kiếm

Nghiên cứu này chủ yếu sử dụng Google Scholar như nguồn tìm kiếm chính để thu thập dữ liệu. Google Scholar cung cấp khả năng tìm kiếm linh hoạt, cho phép truy cập vào bản sao số hoặc bài báo trực tuyến. Sự linh hoạt này không chỉ nhanh chóng thu thập thông tin mà còn giúp đảm bảo tính đa dạng và phong phú của nguồn dữ liệu. Bằng cách này, nghiên cứu đạt được sự toàn diện trong việc đánh giá và tổng hợp thông tin từ các nguồn đa dạng trên Google Scholar.

### 2.2 Tiêu chí tìm kiếm và điều kiện

Tác giả lựa chọn các bài báo để phân tích tổng quan từ cơ sở dữ liệu tạp chí, hội thảo tìm kiếm bởi Google Scholar. Tiêu chí lựa chọn bài báo: i) Thuật ngữ tìm kiếm: ít nhất một thuật ngữ liên quan đến Large Language Models (LLMs), Software Engineering (SE) phải xuất hiện trong tiêu đề bài viết; ii) Thuật ngữ “Large Language Models” và “Software Engineering”.

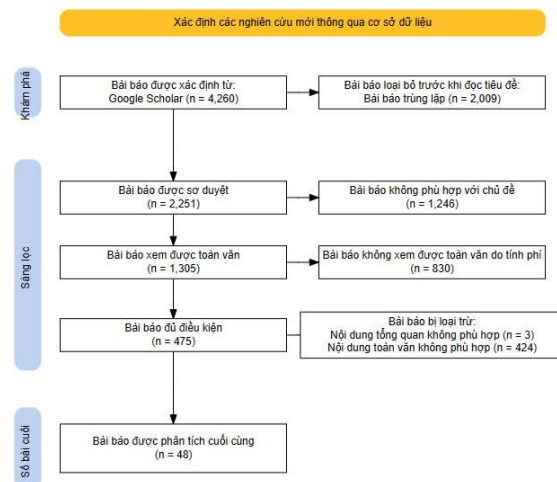
Để được đưa vào danh sách cuối cùng để thực hiện phân tích, đánh giá, các bài báo cần phải đảm bảo một số yêu cầu thêm như sau:

- Thời gian: bài báo được công bố từ 2017 tới nay
  - Ngôn ngữ: bài báo được viết bằng tiếng Anh, tiếng Việt
  - Khả năng truy cập: bài báo có thể truy cập toàn văn
- Những bài báo có một trong các yếu tố sau sẽ bị loại khỏi danh sách:

- Bài báo không viết bằng tiếng Anh hoặc tiếng Việt
- Bài báo xuất bản trước năm 2017
- Bài báo liên quan đến Software Engineering và ngược lại nhưng không liên quan đến LLMs.

H.1 mô tả luồng thông tin qua các giai đoạn khác nhau để đánh giá hệ thống và sử dụng phương pháp PRISMA. Với tổng số 4260 tiêu đề được tìm thấy, loại bỏ lần 1 về tiêu đề 2009 bài; loại bỏ lần 2 không phù hợp với chủ đề 1246 bài; loại bỏ lần 3 không xem được toàn văn 830 bài; loại bỏ lần 4 với nội dung tổng quan không phù hợp (Reason1= 3 bài) và nội dung toàn văn không phù hợp (Reason2= 424 bài).

Cuối cùng, 48 bài đã được đưa vào nghiên cứu này để đánh giá, phân tích.



H.1 Biểu đồ thể hiện sự di chuyển của thông tin qua các giai đoạn khác nhau của một đánh giá hệ thống

## 3. Kết quả và thảo luận

### 3.1 Số lượng bài báo nghiên cứu trong cơ sở dữ liệu tạp chí, hội thảo đối với việc sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm thay đổi như thế nào?

Biểu đồ H.2 cho thấy tổng quan về phân phối số lượng các tài liệu về sử dụng về Large language models trong lĩnh vực Software Engineering từ năm 2017 đến năm 2023. Với tổng số bài được phân tích là 48 với số lượng các bài báo được phân bố không đồng đều, các lĩnh vực này phát triển chậm từ giai đoạn từ 2017 đến 2022 với số lượng liên quan trong cơ sở dữ liệu tìm kiếm là dưới 5 bài, đến năm 2023 các dữ liệu xuất bản liên quan đến từ khóa đột biến tối đa lên đến 38 bài.

Các xu hướng xuất bản hiện nay về thời gian được tìm kiếm từ 2017 đến 2023:



H.2 Số lượng bài báo, sách được xuất bản theo năm từ 2017 đến 2023

**Bảng 1:** Bảng số liệu các bài báo về sử dụng Large language models trong các lĩnh vực về kỹ thuật phần mềm, tự động hóa, kiểm thử, trí tuệ nhân tạo

Nguồn tạp chí	Số lượng bài báo
arXiv preprint [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27]	25
IEEE [28], [29], [30], [31], [32], [33], [34], [35], [36]	9
Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering[37], [38]	2
Design, Automation & Test in Europe Conference & Exhibition [39]	1
AICHe Journal [40]	1
Chi conference on human factors in computing systems extended abstracts [41]	1
Forecasting [42]	1
International Conference on Bridging the Gap between AI and Reality [43]	1
International Conference on Machine Learning [44]	1
Meta-Radiology [45]	1
Nature [46]	1
Proceedings of the 16th ACM SIGPLAN International Conference on Software Language Engineering [47]	1
Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1 [48]	1
Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security [49]	1
Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology[50]	1

Từ **Bảng 1** trên, ta thấy rằng arXiv preprint đóng góp nhiều bài viết nhất có liên quan đến Large language Models, Software Engineering với 25 bài có tỷ lệ 52% bài được phân tích. Đứng thứ 2 là IEEE với 9 bài có tỷ lệ là 19%. Các tạp chí, hội thảo còn lại gần tương đương nhau. Kết quả này cung cấp thông tin hữu ích cho các nhà nghiên cứu về LLMs trong các lĩnh vực về kỹ thuật phần mềm, AI, ChatGPT... Tỷ lệ công bố trên nền tảng Arxiv, một cơ sở dữ liệu điện tử dạng tiền in ấn, không trải qua quá trình phản biện ngang hàng, chiếm 52% có thể được giải thích bởi ba lý do chính. Thứ nhất, các công bố này đều rất mới, chưa được các tạp chí uy tín chấp nhận xuất bản. Thứ hai, các tác giả mong muốn nhận được những nhận định từ cộng đồng trước khi xuất bản chính thức. Cuối cùng, Tuyên bố chủ quyền về lĩnh vực nghiên cứu của các tác giả và nhóm tác giả.

Từ **Bảng 2**, ta có thể thấy rằng bài báo nghiên cứu Evaluating Large Language Models Trained on Code đăng trên arXiv preprint của nhóm tác giả Mark Chen và cộng sự được các nhà khoa học quan tâm và trích dẫn nhiều nhất với số lượt trích dẫn là 1385 trong cơ sở

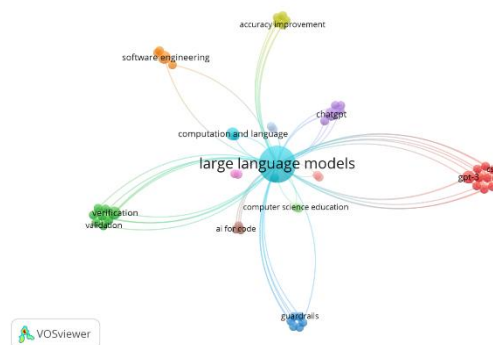
dữ liệu của Google Scholar [24]. Điều này cho thấy vị trí và tầm quan trọng của công trình của họ trong lĩnh vực này. Các bài viết khác nghiên cứu về LLMs trong lĩnh vực kỹ thuật phần mềm cũng được lượng trích dẫn lớn như “Summary of chatgpt-related research and perspective towards the future of large language models”[44]; “Large language models encode clinical knowledge” [45]. Bài viết của các tác giả còn lại có số lượng trích dẫn bằng 0 hoặc 1, cho thấy chưa có nhiều người quan tâm đến bài viết của tác giả.

**Bảng 2.** Top bài báo có số lượt được trích dẫn nhiều nhất trong cơ sở dữ liệu có liên quan đến Large language models

Tên bài báo	Lượt được trích dẫn
Evaluating Large Language Models Trained on Code	1385
Summary of chatgpt-related research and perspective towards the future of large language models	295
Large language models encode clinical knowledge	258
Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models	235
Progprompt: Generating situated robot task plans using large language models	220
Smoothquant: Accurate and efficient post-training quantization for large language models	138
Promptchainer: Chaining large language model prompts through visual programming	107

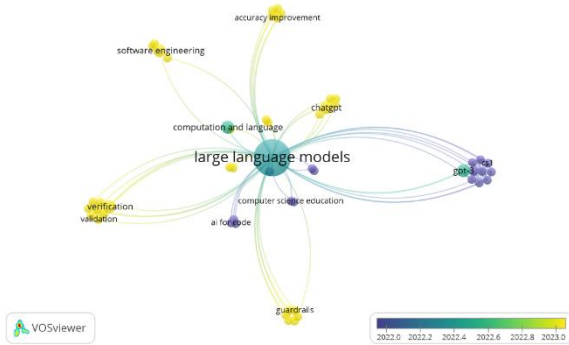
**3.2 Những từ khoá nào xuất hiện thường xuyên nhất về sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm của kho dữ liệu tạp chí, hội thảo?**

**H. 3** cho thấy một bản đồ trực quan hóa các xu hướng nghiên cứu thông qua sử dụng phần mềm Vosviewer về sử dụng Large Language Models trong lĩnh vực kỹ thuật phần mềm từ năm 2017 đến năm 2023. Kết quả lập bản đồ cho thấy các cụm trọng tâm cho các chủ đề nghiên cứu sự ảnh hưởng của sử dụng LLMs đối với Kỹ thuật phần mềm, ChatGPT, cải thiện độ chính xác, ngôn ngữ và tính toán, trí tuệ nhân tạo cho mã, giáo dục khoa học máy tính, chiến lược học tập, GPT3, thăm định, xác minh.



**H. 3** Các lĩnh vực nghiên cứu liên quan đến Large Language Models

**H. 4** thể hiện sự phân bố theo năm các nội dung liên quan đến lĩnh vực nghiên cứu Large Language Models. Trong đó màu vàng thể hiện các nghiên cứu được thực hiện trong năm 2023 có các lĩnh vực về Kỹ thuật phần mềm, ChatGPT, thẩm định, xác minh.. màu tím thể hiện các nghiên cứu được thực hiện trong năm từ 2017 đến 2022 có các lĩnh vực như AI for Code, giáo dục khoa học máy tính...

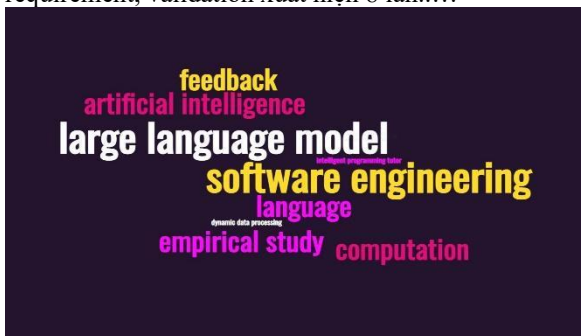


**H. 4** Sự phân bố theo năm các lĩnh vực nghiên cứu liên quan đến Large Language Models

Selected	Keyword	Occurrences	Total link strength
<input checked="" type="checkbox"/>	large language models	17	53
<input checked="" type="checkbox"/>	gpt-3	2	11
<input checked="" type="checkbox"/>	automated feedback	1	10
<input checked="" type="checkbox"/>	code explanations	1	10
<input checked="" type="checkbox"/>	cs1	1	10
<input checked="" type="checkbox"/>	exercise generation	1	10
<input checked="" type="checkbox"/>	natural language generation	1	10
<input checked="" type="checkbox"/>	openai codex	1	10
<input checked="" type="checkbox"/>	programming exercises	1	10
<input checked="" type="checkbox"/>	resource generation	1	10
<input checked="" type="checkbox"/>	robosourcing	1	10
<input checked="" type="checkbox"/>	verification	2	9
<input checked="" type="checkbox"/>	challenges, requirements	1	8
<input checked="" type="checkbox"/>	design	1	8
<input checked="" type="checkbox"/>	gpt: bard	1	8
<input checked="" type="checkbox"/>	large language models- assisted	1	8
<input checked="" type="checkbox"/>	software engineering	1	8
<input checked="" type="checkbox"/>	state-of-the-art	1	8
<input checked="" type="checkbox"/>	validation	1	8
<input checked="" type="checkbox"/>	chatgpt	2	8

**H. 5** Sự phân bố theo năm các lĩnh vực nghiên cứu liên quan đến Large Language Models

**H. 5** và **H. 6** thể hiện những từ khóa xuất hiện thường xuyên nhất như Large Language Models với tổng số 53 lần; GPT-3 với 11 lần; Automated Feedback, code explanations, cs1, exercise generation, openai codex, programming exercises, resources generation, robosourcing xuất hiện 10 lần; Verification xuất hiện 9 lần; Software Engineering, Design, challenges, requirement, validation xuất hiện 8 lần....

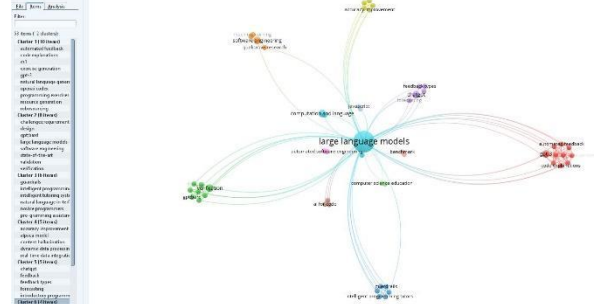


**H. 6** Đám mây từ khóa được trích xuất từ các nội dung tóm tắt của các bài báo liên quan đến Large Language Models

**3.3 Những chủ đề nghiên cứu quan trọng nhất trong việc sử dụng LLMs trong lĩnh vực kỹ thuật phần mềm là gì?**

**H. 7** nhấn mạnh việc tổng hợp các từ khóa đã được sử dụng trong các nghiên cứu giúp cho người đọc hiểu rõ hơn về những mối quan hệ giữa các từ khóa này. Từ

bảng phân tích có các cụm tập trung nội dung Large Language Models về các lĩnh vực Kỹ thuật phần mềm, ChatGPT, thẩm định, xác minh... như sau: Cụm số 1 (Cluster 1) với 10 mục, cụm số 2 (Cluster 2) với 8 mục, cụm số 3 (cluster 3) với 6 mục,... Từ khóa "Large language models" có tần suất xuất hiện cao nhất với 53 lần, từ khóa "Software Engineering" đứng thứ 4 với xuất hiện 8 lần...



**H. 7** Phân tích sự xuất hiện của từ khóa

Những chủ đề này có thể là những hướng nghiên cứu quan trọng trong tương lai để tối ưu hóa sức mạnh của Large Language Models trong lĩnh vực Kỹ thuật phần mềm.

**3.4 Những khoảng trống và lĩnh vực cho nghiên cứu trong tương lai là gì?**

Large language models đã được nghiên cứu ở mức độ trung bình từ 2017 đến 2022 nhưng đến năm 2023 đã dần dần phát triển mạnh, nhưng vẫn còn một số hạn chế và thách thức cần phải vượt qua. **Bảng 3** chỉ ra tần suất các hạn chế, thách thức của Large language models được nêu trong các ấn phẩm.

**Bảng 3.** Tần suất các hạn chế được chỉ ra trong các ấn phẩm

Hạn chế / thách thức	Tài liệu tham khảo
Không nắm bắt được một số bảo mật nhất định	[51], [52], [53], [54]
Không khái quát hóa và lập trình các ngôn ngữ khác	[51]
Dung lượng bộ nhớ	[55], [56]
Đánh giá chất lượng	[52], [53]
Nội dung sai lệch	[54], [55], [56]
Phương pháp tiếp cận	[57], [58], [59]
Tạo ra hành vi phức tạp của robot vẫn còn tương đối ít được khám phá.	[60]
Các vấn đề liên quan xây dựng điểm chuẩn, thiết kế phương pháp và mở rộng chức năng	[61]
Mối quan tâm về khả năng tái sản xuất	[50]
Bộ dữ liệu khác	[62]
Phản ứng thái quá trước mỗi đe dọa tiềm tàng của công nghệ	[63]

Thiếu dữ liệu	[54]
Tự động hóa xử lý dữ liệu, quy trình	[53], [64]
Kiểm thử, kiểm chứng, kiểm tra dữ liệu	[65]

#### 4. Kết luận

Ở bài viết này, nghiên cứu về Large Language Models (LLMs) trong các lĩnh vực kỹ thuật phần mềm với mục đích phân tích, đánh giá các bài báo, tài liệu liên quan trong việc sử dụng LLMs từ năm 2017 đến năm 2023. Các kết quả nghiên cứu cho thấy giai đoạn từ năm 2017 đến 2022 có ít số lượng nghiên cứu về LLMs ở mức ít 45% và tăng vọt lên trong năm 2023 với 55% trong các lĩnh vực kỹ thuật phần mềm, kiểm thử, ChatGPT, kiểm định... Các xuất bản tập đầu là "arXiv preprint", tiếp đến là "IEEE" trong việc nghiên cứu về LLMs và bài viết "Evaluating Large Language Models Trained on Code" là nghiên cứu LLMs được trích dẫn 1385 lần có ảnh hưởng lớn nhất trong lĩnh vực kỹ thuật phần mềm. Trong các nghiên cứu LLMs về các lĩnh vực kỹ thuật phần mềm, ChatGPT, kiểm định, giáo dục... thì quốc gia tập trung nhiều xuất bản, hội thảo nhất là Hoa Kỳ. Tuy nhiên, việc đánh giá tổng quan nghiên cứu LLMs trong các lĩnh vực kỹ thuật phần mềm, ChatGPT... mà các tác giả chưa đề cập cũng như phân tích nhiều. Qua bài đánh giá này làm căn cứ xác định khoảng trống nghiên cứu trong tương lai cũng như định hướng thêm trong việc phát triển các nghiên cứu của LLMs với lĩnh vực kỹ thuật phần mềm.

#### Tài liệu tham khảo

[1] M. Chen *et al.*, "Evaluating large language models trained on code," *arXiv preprint arXiv:2107.03374*, 2021.

[2] P. Vaithilingam, T. Zhang, and E. L. Glassman, "Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models," in *Chi conference on human factors in computing systems extended abstracts*, 2022, pp. 1–7.

[3] J. Li, G. Li, C. Tao, H. Zhang, F. Liu, and Z. Jin, "Large Language Model-Aware In-Context Learning for Code Generation," *arXiv preprint arXiv:2310.09748*, 2023.

[4] Z. Zhang, X. Zhang, W. Xie, and Y. Lu, "Responsible Task Automation: Empowering Large Language Models as Responsible Task Automators," *arXiv preprint arXiv:2306.01242*, 2023.

[5] J. Chanenson, M. Pickering, and N. Apthorpe, "Automating Governing Knowledge Commons and Contextual Integrity (GKC-CI) Privacy Policy Annotations with Large Language Models," *arXiv preprint arXiv:2311.02192*, 2023.

[6] A. Robinson *et al.*, "Bio-sieve: Exploring instruction tuning large language models for systematic review automation," *arXiv preprint arXiv:2308.06610*, 2023.

[7] C. Tang *et al.*, "PolicyGPT: Automated Analysis of Privacy Policies with Large Language

Models. *arXiv 2023*," *arXiv preprint cs.CL/2309.10238*.

[8] Z. Cao, Y. Yang, and H. Zhao, "Autohall: Automated hallucination dataset generation for large language models," *arXiv preprint arXiv:2310.00259*, 2023.

[9] Y. Shen *et al.*, "TaskBench: Benchmarking Large Language Models for Task Automation," *arXiv preprint arXiv:2311.18760*, 2023.

[10] D. Noever, "Can large language models find and fix vulnerable software?," *arXiv preprint arXiv:2308.10345*, 2023.

[11] Q. Zhang *et al.*, "A critical review of large language model on software engineering: An example from chatgpt and automated program repair," *arXiv preprint arXiv:2310.08879*, 2023.

[12] S. MacNeil *et al.*, "Automatically Generating CS Learning Materials with Large Language Models," *arXiv preprint arXiv:2212.05113*, 2022.

[13] J. Wang, Y. Huang, C. Chen, Z. Liu, S. Wang, and Q. Wang, "Software testing with large language model: Survey, landscape, and vision," *arXiv preprint arXiv:2307.07221*, 2023.

[14] J. Shin, C. Tang, T. Mohati, M. Nayebi, S. Wang, and H. Hemmati, "Prompt Engineering or Fine Tuning: An Empirical Assessment of Large Language Models in Automated Software Engineering Tasks," *arXiv preprint arXiv:2310.10508*, 2023.

[15] X. Du *et al.*, "Classeval: A manually-crafted benchmark for evaluating llms on class-level code generation," *arXiv preprint arXiv:2308.01861*, 2023.

[16] G. Deng *et al.*, "Masterkey: Automated jailbreak across multiple large language model chatbots," *arXiv preprint arXiv:2307.08715*, 2023.

[17] S. Thakur *et al.*, "Verigen: A large language model for verilog code generation," *arXiv preprint arXiv:2308.00708*, 2023.

[18] J. Z. Pan *et al.*, "Large language models and knowledge graphs: Opportunities and challenges," *arXiv preprint arXiv:2308.06374*, 2023.

[19] R. Puri *et al.*, "Codenet: A large-scale ai for code dataset for learning a diversity of coding tasks," *arXiv preprint arXiv:2105.12655*, 2021.

[20] M. Liffiton, B. Sheese, J. Savelka, and P. Denny, "Codehelp: Using large language models with guardrails for scalable support in programming classes," *arXiv preprint arXiv:2308.06921*, 2023.

[21] Z. Zheng *et al.*, "Towards an understanding of large language models in software engineering tasks," *arXiv preprint arXiv:2308.11396*, 2023.

[22] J. Wei *et al.*, "Emergent abilities of large language models," *arXiv preprint arXiv:2206.07682*, 2022.

[23] L. Weidinger *et al.*, "Ethical and social risks of harm from language models (2021)," *arXiv preprint arXiv:2112.04359*, 2021.

[24] M. Nadeem, A. Bethke, and S. Reddy, "StereoSet: Measuring stereotypical bias in pretrained language models," *arXiv preprint arXiv:2004.09456*, 2020.



- [25] M. Chen *et al.*, “Evaluating large language models trained on code,” *arXiv preprint arXiv:2107.03374*, 2021.
- [26] I. Beltagy, K. Lo, and A. Cohan, “SciBERT: A pretrained language model for scientific text,” *arXiv preprint arXiv:1903.10676*, 2019.
- [27] H. K. Dam, T. Tran, and T. Pham, “A deep language model for software code,” *arXiv preprint arXiv:1608.02715*, 2016.
- [28] K. Hanifi, O. Cetin, and C. Yilmaz, “On ChatGPT: Perspectives from Software Engineering Students,” in *2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security (QRS)*, IEEE, 2023, pp. 196–205.
- [29] Y. Fu *et al.*, “Gpt4aigchip: Towards next-generation ai accelerator design automation via large language models,” in *2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, IEEE, 2023, pp. 1–9.
- [30] Z. Fan, X. Gao, M. Mirchev, A. Roychoudhury, and S. H. Tan, “Automated repair of programs from large language models,” in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, IEEE, 2023, pp. 1469–1481.
- [31] S. Kang, J. Yoon, and S. Yoo, “Large language models are few-shot testers: Exploring llm-based general bug reproduction,” in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, IEEE, 2023, pp. 2312–2323.
- [32] I. Singh *et al.*, “Progprompt: Generating situated robot task plans using large language models,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 11523–11530.
- [33] N. Kiesler, D. Lohr, and H. Keuning, “Exploring the potential of large language models to generate formative programming feedback,” in *2023 IEEE Frontiers in Education Conference (FIE)*, IEEE, 2023, pp. 1–5.
- [34] J. Lu, L. Yu, X. Li, L. Yang, and C. Zuo, “LLaMA-Reviewer: Advancing Code Review Automation with Large Language Models through Parameter-Efficient Fine-Tuning,” in *2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*, IEEE, 2023, pp. 647–658.
- [35] I. Ozkaya, “Application of Large Language Models to Software Engineering Tasks: Opportunities, Risks, and Implications,” *IEEE Softw*, vol. 40, no. 3, pp. 4–8, 2023.
- [36] M. Schäfer, S. Nadi, A. Eghbali, and F. Tip, “An empirical evaluation of using large language models for automated unit test generation,” *IEEE Transactions on Software Engineering*, 2023.
- [37] E. First, M. Rabe, T. Ringer, and Y. Brun, “Baldur: Whole-proof generation and repair with large language models,” in *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2023, pp. 1229–1241.
- [38] J. Zhang, P. Nie, J. J. Li, and M. Gligoric, “Multilingual code co-evolution using large language models,” in *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2023, pp. 695–707.
- [39] S. Thakur *et al.*, “Benchmarking Large Language Models for Automated Verilog RTL Code Generation,” in *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE, 2023, pp. 1–6.
- [40] E. Hirtreiter, L. Schulze Balhorn, and A. M. Schweidtmann, “Toward automatic generation of control structures for process flow diagrams with large language models,” *AIChE Journal*, p. e18259, 2023.
- [41] P. Vaithilingam, T. Zhang, and E. L. Glassman, “Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models,” in *Chi conference on human factors in computing systems extended abstracts*, 2022, pp. 1–7.
- [42] S. Makridakis, F. Petropoulos, and Y. Kang, “Large language models: Their success and impact,” *Forecasting*, vol. 5, no. 3, pp. 536–549, 2023.
- [43] L. Belzner, T. Gabor, and M. Wirsing, “Large language model assisted software engineering: prospects, challenges, and a case study,” in *International Conference on Bridging the Gap between AI and Reality*, Springer, 2023, pp. 355–374.
- [44] G. Xiao, J. Lin, M. Seznec, H. Wu, J. Demouth, and S. Han, “Smoothquant: Accurate and efficient post-training quantization for large language models,” in *International Conference on Machine Learning*, PMLR, 2023, pp. 38087–38099.
- [45] Y. Liu *et al.*, “Summary of chatgpt-related research and perspective towards the future of large language models,” *Meta-Radiology*, p. 100017, 2023.
- [46] K. Singhal *et al.*, “Large language models encode clinical knowledge,” *Nature*, vol. 620, no. 7972, pp. 172–180, 2023.
- [47] F. Ribeiro, J. N. C. de Macedo, K. Tsushima, R. Abreu, and J. Saraiva, “GPT-3-Powered Type Error Debugging: Investigating the Use of Large Language Models for Code Repair,” in *Proceedings of the 16th ACM SIGPLAN International Conference on Software Language Engineering*, 2023, pp. 111–124.
- [48] S. Sarsa, P. Denny, A. Hellas, and J. Leinonen, “Automatic generation of programming exercises and code explanations using large language models,” in *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1*, 2022, pp. 27–43.
- [49] J. He and M. Vechev, “Large language models for code: Security hardening and adversarial testing,” in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2023, pp. 1865–1879.
- [50] T. S. Kim, Y. Lee, M. Chang, and J. Kim, “Cells, generators, and lenses: Design framework for object-oriented interaction with large language models,” in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023, pp. 1–18.

[51] J. Wei *et al.*, “Emergent abilities of large language models,” *arXiv preprint arXiv:2206.07682*, 2022.

[52] J. Sallou, T. Durieux, and A. Panichella, “Breaking the Silence: the Threats of Using LLMs in Software Engineering,” *arXiv preprint arXiv:2312.08055*, 2023.

[53] N. Kiesler, D. Lohr, and H. Keuning, “Exploring the potential of large language models to generate formative programming feedback,” in *2023 IEEE Frontiers in Education Conference (FIE)*, IEEE, 2023, pp. 1–5.

[54] Y. Liu *et al.*, “Summary of chatgpt-related research and perspective towards the future of large language models,” *Meta-Radiology*, p. 100017, 2023.

[55] G. Xiao, J. Lin, M. Seznec, H. Wu, J. Demouth, and S. Han, “Smoothquant: Accurate and efficient post-training quantization for large language models,” in *International Conference on Machine Learning*, PMLR, 2023, pp. 38087–38099.

[56] Z. Zheng *et al.*, “Towards an understanding of large language models in software engineering tasks,” *arXiv preprint arXiv:2308.11396*, 2023.

[57] Z. Fan, X. Gao, M. Mirchev, A. Roychoudhury, and S. H. Tan, “Automated repair of programs from large language models,” in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, IEEE, 2023, pp. 1469–1481.

[58] J. Z. Pan *et al.*, “Large language models and knowledge graphs: Opportunities and challenges,” *arXiv preprint arXiv:2308.06374*, 2023.

[59] S. Thakur *et al.*, “Verigen: A large language model for verilog code generation,” *arXiv preprint arXiv:2308.00708*, 2023.

[60] X. Du *et al.*, “Classeval: A manually-crafted benchmark for evaluating llms on class-level code generation,” *arXiv preprint arXiv:2308.01861*, 2023.

[61] I. Singh *et al.*, “Progprompt: Generating situated robot task plans using large language models,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 11523–11530.

[62] Z. Zhang, X. Zhang, W. Xie, and Y. Lu, “Responsible Task Automation: Empowering Large Language Models as Responsible Task Automators,” *arXiv preprint arXiv:2306.01242*, 2023.

[63] J. Li, G. Li, C. Tao, H. Zhang, F. Liu, and Z. Jin, “Large Language Model-Aware In-Context Learning for Code Generation,” *arXiv preprint arXiv:2310.09748*, 2023.

[64] S. Makridakis, F. Petropoulos, and Y. Kang, “Large language models: Their success and impact,” *Forecasting*, vol. 5, no. 3, pp. 536–549, 2023.

[65] Y. Fu *et al.*, “Gpt4aigchip: Towards next-generation ai accelerator design automation via large language models,” in *2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, IEEE, 2023, pp. 1–9.

[66] J. Lu, L. Yu, X. Li, L. Yang, and C. Zuo, “LLaMA-Reviewer: Advancing Code Review Automation with Large Language Models through Parameter-Efficient Fine-Tuning,” in *2023 IEEE 34th*

*International Symposium on Software Reliability Engineering (ISSRE)*, IEEE, 2023, pp. 647–658.



**Nguyễn Văn Việt**, sinh năm 1986, nhận bằng Thạc sĩ về Công nghệ thông tin của trường ĐH tổng hợp Enverga, Philippines năm 2012. Hiện tại là giảng viên bộ môn Công nghệ phần mềm- trường ĐH Công nghệ thông tin và truyền thông – ĐH Thái Nguyên. Từ năm 2023 là nghiên cứu sinh Trường ĐH Công nghệ tin và truyền thông – ĐH Thái Nguyên, hướng nghiên cứu chính là Khoa học máy tính, công nghệ phần mềm.



**Nguyễn Hữu Khánh**, sinh năm 1987, nhận bằng Thạc sĩ về Khoa học máy tính của trường Đại học Công nghệ thông tin và truyền thông – Đại học Thái Nguyên năm 2022. Hiện tại là chuyên viên Trung tâm Đào tạo từ xa - ĐH Thái nguyên. Từ năm 2023 là nghiên cứu sinh trường ĐH Công nghệ thông tin và Truyền thông – ĐHTN, hướng nghiên cứu chính là xử lý ngôn ngữ tự nhiên và thị giác máy tính.



**Nguyễn Thị Duyên**, sinh năm 1987, nhận bằng Thạc sĩ về Công nghệ thông tin của trường ĐH tổng hợp Enverga, Philippines năm 2014. Hiện tại là giảng viên bộ môn Mạng và An toàn thông tin Trường Đại học Công nghệ thông tin và truyền thông – ĐH Thái Nguyên. Từ năm 2023 là nghiên cứu sinh Trường ĐH Công nghệ tin và truyền thông – ĐH Thái Nguyên. Hướng nghiên cứu chính là Khoa học máy tính.



**Lương Thị Minh Huế**, sinh năm 1986, nhận bằng Thạc sĩ về Công nghệ thông tin của trường ĐH tổng hợp Enverga, Philippines năm 2013. Hiện tại là giảng viên bộ môn Mạng và An toàn thông tin Trường Đại học Công nghệ thông tin và truyền thông – ĐH Thái Nguyên. Hướng nghiên cứu chính là Khoa học máy tính, truyền thông và mạng máy tính.



**Nguyễn Thế Vĩnh**, sinh năm 1985, nhận bằng Tiến sĩ về Khoa học máy tính của Trường Đại học kỹ thuật Texas năm 2020. Hiện tại là giảng viên ngành Kỹ thuật phần mềm của Trường Đại học Công nghệ thông tin và truyền thông - ĐH Thái Nguyên. Hướng nghiên cứu chính là Khoa học máy tính, AI.





**Nguyễn Hữu Công**, sinh năm 1964, nhận bằng Tiến sĩ về điều khiển tự động của trường ĐH Bách khoa Hà Nội năm 2003, bổ nhiệm Phó giáo sư năm 2007. Hiện tại là Phó Giám đốc Đại học Thái nguyên. Hướng nghiên cứu chính là điều khiển tự động, điều khiển tối ưu cho các đối tượng có tham số phân bố, biến đổi chậm.