The First International Workshop on Large Language Model-Oriented Empirical Research (LLanMER)

Na Meng nm8247@vt.edu Virginia Tech Blacksburg, Virginia, USA Xiaoyin Wang The University of Texas at San Antonio San Antonio, Texas, USA xiaoyin.wang@utsa.edu Chris Brown Virginia Tech Blacksburg, Virginia, USA dcbrown@vt.edu

ABSTRACT

Ever since the launch of ChatGPT in 2022, there have been fast growing interests in the application of ChatGPT and other large language models (LLMs). Researchers have conducted various studies or created tools to integrate LLMs into (1) the current practices of software development as well as maintenance, (2) the training of next-generation software engineers, and (3) human-computer interactions to facilitate better software/hardware usage. However, lots of research questions remain open regarding the methodologies for conducting empirical research with LLMs. For instance, what is the best usage of LLMs in different scenarios, what are rigorous measurements for LLM results, and what are the potential impacts of LLM-oriented research on ethics, economy, energy, as well as environment? All these questions are critical for researchers to conduct responsible, reliable, and reproducible empirical research with LLMs. Thus, we propose to organize a new workshop focusing on methodologies for conducting empirical research with LLMs. This workshop intends to provide a venue for researchers and practitioners to share ideas, discuss obstacles and challenges in LLM-oriented empirical research, brainstorm solutions, and establish collaborations to define reliable LLM-oriented empirical methodologies in cost-efficient ways. To achieve that goal, our workshop will include a keynote talk, paper presentations, and a panel.

KEYWORDS

LLM, empirical research, software engineering

ACM Reference Format:

1 WORKSHOP FORMAT

Our workshop will include a keynote presentation, two sessions of paper presentations, and a panel. Currently we plan to have all authors attend in person, but due to different policies in different

Conference'17, July 2017, Washington, DC, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00 https://doi.org/XXXXXXXXXXXXXXX countries/states, we may also have the workshop in a hybrid format and allow some presentations to be delivered online.

The intended length of the workshop is one day, and the preferred date is the day right after the main conference. The preliminary schedule is as below. The concrete time will be adjusted based on FSE program and the number of papers accepted.

- 9:15 9:30 Opening
- 9:30 10:30 Keynote
- 10:30 11:00 Break
- 11:00 12:30 Paper Session I
- 12:30 14:00 Lunch
- 14:00 15:30 Paper Session II
- 15:30 16:00 Break
- 16:00 17:00 Panel
- 17:00 17:15 Closing

To generate and stimulate discussion at the workshop, before the workshop, we will send invitation emails to researchers and groups that conduct LLM-related research. In those emails, we will invite people to submit papers to our workshop, give talks, and/or contribute discussion questions that they are interested about. During the workshop, we will organize a panel to include panelists that have relevant experience from either industry or academia, and present a list of discussion questions based on people's interests, and the workshop's focus on empirical methodologies.

The facility required by our workshop is similar to other workshops. For the in-person part, we need a room with a capacity of 30-40 and a projector or a large screen. For the online part, we need a virtual room. **The workshop proceedings will appear in the ACM digital library.**

2 PAPER SELECTION PROCEDURE

Types of Contributions In our Call for Paper (CFP), we will call for two types of submissions to the workshop:

- *Long* (up to 6 pages), or *Short* (up to 3 pages) Research Papers, plus 1-2 pages for references and well-marked appendices. These papers present research work in the early stage. Position papers with exceptional visions will also be considered.
- Long (up to 6 pages), or Short (up to 3 pages) Experience Papers, plus 1-2 pages for references and well-marked appendices. These submissions should report experience of applying or assessing LLMs in a non-trivial setting.

Review and Evaluation Proceess Before sending out the CFP mentioned above, we will formulate a program committee of about

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

15 people. To construct this committee, we will reach out to companies/organizations that have LLM products (e.g., Microsoft, OpenAI, Google, and Meta), and academic research groups that conduct research with LLMs. We will endeavor to ensure the diversity of our program committee, by including people with different colors, genders, and backgrounds.

To ensure fairness, we will follow a double-blind reviewing process. Submitted papers must (a) omit any reference to the authors' names or the names of their institutions, and (b) reference the authors' own related work in the third person (e.g., not "We build on our previous work ..." but rather "We build on the work of ..."). Any paper violating the double-blind policy will be desk-rejected. Each paper will be reviewed by three people. After all reviewers submitted their reviews, the PC chairs will lead a discussion for any paper receiving divergent reviews (i.e., with both "accept" and "reject" scores), until the reviewers come to a consensus.

Important Dates The expected dates for paper submission and notification of acceptance are separately February 25th, 2025 and March 25, 2025.

Program Committee We have not formulated the entire committee. Dr. Na Meng and Dr. Xiaoyin Wang will both serve on the committee. We also have a tentative list of candidate reviewers, and will reach out to people once the proposal is accepted.

Online Submission System We will use Easychair as our online submission system.

Website Link: The preliminary website of our workshop is available at https://people.cs.vt.edu/nm8247/llanmer/.

Proceeding Chair Dr. Chris Brown will be responsible for sending the list of accepted papers to the publishers.

3 PUBLICITY PLANS

The target audience of our workshop are researchers, practitioners, and students who are interested in (a) the empirical research with LLMs, or (b) the integration of LLM usage in software/hardware products. The authors already have a list of candidate speakers to invite, and a list of research groups that may submit to the workshop. We expect 30-40 attendants to participate in our workshop.

The workshop will be open to everybody. To solicit and select workshop participants/presenters, we will publicize our workshop with the following mechanisms:

- Targeting known companies or organizations working on LLMs. We will search for owners or maintainers of opensource and close-source LLMs, and invite them to submit papers to describe their LLM experience or new software products. We already identified several companies like Microsoft, OpenAI, Google, and Meta.
- Targeting known research groups in academia. We will search Google Scholar for research papers or technical reports related to LLMs. We will contact the authors to check whether they want to submit papers to report the recent research findings. We already identified several research groups from University of Illinois Urbana-Champaign, Virginia Tech, University of Texas at San Antonio, and Peking University (China).

- Taking advantage of community mailing lists. We will send out call-for-papers to major mailing lists like SEWorld. We will also share the information through social network platforms such as Facebook, X, and WeChat.
- Personal Relations. The chairs and PC members will also send call-for-papers to their personal networks, inviting people to submit papers and spread the words.

4 WORKSHOP HISTORY

This is the first time we will organize the new workshop, so there is no history data of previous events.

5 ORGANIZER INFORMATION

Na Meng is an associate professor at Virginia Tech. She has been studying LLM application [2], program analysis [18], program debugging [4, 6], automatic testing [16], the application of AI techniques to predict developers' decision making for software maintenance activities [1], and vulnerability detection as well as repair [21]. Na served on the organization committee (OC) of IEEE Secure Development Conference 2022 (SecDev 2022) as a vise general chair. She served on the OC of SecDev 2023 as the general chair. Na is serving on the steering committee of SecDev as a committee member. SecDev is an independent security conference, having 61–68 attendees each year during 2022-2023. Na also served on the program or organization committees of PLDI 2019-2021, ASE 2017-2018, ICSE 2019-2021/2024, and ICSME 2017/2019.

Xiaoyin Wang is an associate professor at the University of Texas at San Antonio. He has been working in the area of AIbased software engineering, including validation of LLM-generated code [17], repairing build script by learning historical data [7], inferring data types by learning from runtime data [20], detecting security-related icons and corresponding information leaks [19], predicting test oracles based on user interface screenshots [9], recommending copy-and-paste decisions [15], mining and predicting security and duplicate bugs reports [8], etc. Xiaoyin served as the proceedings chair of ICSE 2019, the organization chair of SoftMine (Workshop on Software Mining) 2015-2018, and the organization chair of VARSE (Virtual and Augmented Reality Software Engineering) 2022 and 2024 (co-located with ASE).

Chris Brown is an assistant professor at Virginia Tech. He has explored capabilities and opportunities for LLMs to promote evidence-based beliefs and behaviors in SE [3], support Computer Science and SE education [13, 14], inform hiring decisions by facilitating technical interview preparation [11] and conducting resume matching tasks [12], impact information-seeking in SE [5], and to generate human-centric dynamic analysis security tool reports [10]. Chris also has experience served on the organization committee of FSE 2024 and VI/HCC 2023. He is the workshop co-chair for Visual Languages and Human-Centric Computing (VL/HCC) 2025, and an organizer of the proposed next iteration of the Software Developer Diversity and Inclusion (SDDI) workshop.

6 CALL FOR SUBMISSIONS

The one page call-for-papers is on the last page of this proposal.

The First International Workshop on Large Language Model-Oriented Empirical Research (LLanMER)

REFERENCES

- W. Aldndni, N. Meng, and F. Servant. Automatic prediction of developers' resolutions for software merge conflicts. *Journal of Systems and Software*, 206:111836, 2023.
- [2] S. M. H. Anik, X. Gao, and N. Meng. Towards automated occupant profile creation in smart buildings: A machine learning-enabled approach for user persona generation. *Energy and Buildings*, 297:113485, 2023.
- [3] C. Brown and J. Cusati. Exploring the evidence-based beliefs and behaviors of llm-based programming assistants. arXiv preprint arXiv:2407.13900, 2024. In Submission to: ACM Transactions on Software Engineering and Methodology (TOSEM) New Frontiers in Software Engineering.
- [4] T. Dao, L. Zhang, and N. Meng. How does execution information help with information-retrieval based bug localization? In 2017 IEEE/ACM 25th International Conference on Program Comprehension (ICPC), pages 241–250, 2017.
- [5] E. A. Haque, C. Brown, T. D. LaToza, and B. Johnson. Information seeking using ai assistants, 2024.
- [6] F. Hassan, N. Meng, and X. Wang. Uniloc: Unified fault localization of continuous integration failures. ACM Trans. Softw. Eng. Methodol., 32(6), sep 2023.
- [7] F. Hassan and X. Wang. Hirebuild: An automatic approach to history-driven repair of build scripts. In Proceedings of the 40th international conference on software engineering, pages 1078-1089, 2018.
- [8] S. Mostafa, B. Findley, N. Meng, and X. Wang. Sais: Self-adaptive identification of security bug reports. *IEEE Transactions on Dependable and Secure Computing*, 18(4):1779–1792, 2019.
- [9] T. Rafi, X. Zhang, and X. Wang. Predart: Towards automatic oracle prediction of object placements in augmented reality testing. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, pages 1–13, 2022.
- [10] A. Thool and C. Brown. Harnessing the power of llms to simplify security: Llm summarization for human-centric dast reports. In Visual Languages and Human-Centric Computing, 2024.
- [11] S. Vaishampayan and C. Brown. Will you trust me more than chatgpt? evaluating user perceptions of llm-generated feedback for technical interviews. In In Submission to: International Conference on Software Engineering (ICSE), 2025.
- [12] S. Vaishampayan, HunterLeary, Y. B. Alebachew, L. Hickman, B. A. Stevenor, FletcherWimbush, W. Beck, and C. Brown. Human and Ilm-based resume matching: An observational study. In *In Submission to: ACM Conference on Computer* Supported Cooperative Work & Social Computing (CSCW), 2025.
- [13] T. Wang, D. V. Díaz, C. Brown, and Y. Chen. Exploring the role of ai assistants in computer science education: Methods, implications, and instructor perspectives. In 2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pages 92–102. IEEE, 2023.
- [14] T. Wang, R. Ramanujan, Y. Lu, C. Mao, Y. Chen, and C. Brown. Devcoach: Supporting students in learning the software development life cycle at scale with generative agents. In *Learning at Scale*, 2024.
- [15] X. Wang, Y. Dang, L. Zhang, D. Zhang, E. Lan, and H. Mei. Can i clone this piece of code here? In Proceedings of the 27th IEEE/ACM international conference on automated software engineering, pages 170–179, 2012.
- [16] X. Wang, T. Rafi, and N. Meng. Vrguide: Efficient testing of virtual reality scenes via dynamic cut coverage. In 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), pages 951–962, Los Alamitos, CA, USA, sep 2023. IEEE Computer Society.
- [17] X. Wang and D. Zhu. Validating LLM-Generated Programs with Metamorphic Prompt Testing. https://arxiv.org/abs/2406.06864, 2024.
- [18] Y. Wang, N. Meng, and H. Zhong. An empirical study of multi-entity changes in real bug fixes. In 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME), pages 287–298, 2018.
- [19] X. Xiao, X. Wang, Z. Cao, H. Wang, and P. Gao. Iconintent: automatic identification of sensitive ui widgets based on icon classification for android apps. In 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE), pages 257–268. IEEE, 2019.
- [20] X. Zhang, J. Heaps, R. Slavin, J. Niu, T. Breaux, and X. Wang. Daisy: Dynamicanalysis-induced source discovery for sensitive data. ACM Transactions on Software Engineering and Methodology, 32(4):1–34, 2023.
- [21] Y. Zhang, Y. Xiao, M. M. A. Kabir, D. D. Yao, and N. Meng. Example-based vulnerability detection and repair in java code. In *Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension*, ICPC '22, page 190–201, New York, NY, USA, 2022. Association for Computing Machinery.

Call For Submissions

The First International Workshop on Large Language Model-Oriented Empirical Research (LLanMER) is a research forum for software engineers, researchers, and educators to exchange ideas and build collaborations. The workshop invites papers describing novel ideas and early-stage experiment results concerning LLM-oriented empirical research.

Solicited topics include, but not limited to:

- (1) Methodologies: How should we leverage LLMs to solve real-world problems? In the problem-solution procedure, how can we reveal, measure, and address the hallucination issues of LLMs?
- (2) Measurements: How do we precisely measure the effectiveness of LLM usage and evaluate results? How can we ensure the reproducibility and representativeness of evaluation results? How can we evaluate LLM-based approaches in a scalable way?
- (3) Analytics: How do we compare different usage of LLMs? How do we ensure a fair comparison given the existence of randomness and issues such as hallucination as well as lack of ground truth?
- (4) Ethical Aspect: What approaches can we take to ensure that the LLM-oriented empirical research does not violate ethical regulations or raise ethical concerns?
- (5) Economy Aspect: What is the cost comparison between different usage of LLMs? How are LLM-based approaches compared with non-LLM-based approaches in terms of effectiveness, performance, runtime cost, and financial cost?
- (6) Energy Aspect: What is the energy consumption of different usage of LLMs? What kinds of LLM-oriented approaches are energysaving solutions or energy-consuming solutions? How can we optimize the energy consumption by distinct LLM usage without compromising the effectiveness significantly?
- (7) Environment Aspect: What potential impacts LLM usage can introduce to our environment or society? How does that impact personal privacy, intellectual property, and copyright?

We call for two types of submissions to the workshop:

- *Long* (up to 6 pages), or *Short* (up to 3 pages) **Research Papers**, plus 1-2 pages for references and well-marked appendices. These papers present research work in early stage. Position papers with exceptional visions will also be considered.
- *Long* (up to 6 pages), or *Short* (up to 3 pages) *Experience Papers*, plus 1-2 pages for references and well-marked appendices. These submissions should report experience of applying or assessing LLMs in a non-trivial setting.

All papers will be reviewed by three reviewers based on their novelty and soundness.

Double Anonymous Review

Similar to FSE, LLanMER will employ a double-anonymous review process. Thus, no submission may reveal its authors' identities. The authors must make every effort to honor the double-anonymous review process. In particular:

- Authors' names must be omitted from the submission.
- All references to the author's prior work should be in the third person.
- While authors have the right to upload preprints on ArXiV or similar sites, they must avoid specifying that the manuscript was submitted to LLanMER.
- All communication with the program committee must go through the program committee chairs. Do not contact individual program committee members regarding your submission.

Formatting and Submission Guidelines

Abstracts and papers must be submitted electronically through the LLanMER paper submission website. We will follow the schedule of FSE on our deadlines, notification dates, and other important dates. Authors are required to upload their paper in the PDF format. It should conform, at time of submission, to the IEEE conference proceedings template: https://www.ieee.org/conferences/publishing/templates.html (title in 24pt font and full text in 10pt type, LaTeX users must use \documentclass[10pt, conference]{IEEEtran} without including the compsoc or compsocconf options).

Accepted Papers

After acceptance, the list of paper authors cannot be changed under any circumstances. That is, the list of authors on camera-ready papers must be identical to those on submitted papers. Paper titles cannot be changed except by permission of the chairs and only when reviewers recommended a change for clarity or accuracy with the paper content. If a submission is accepted, at least one author of the paper is required to register for the workshop and present the paper online or in-person. As a published ACM author, you and your co-authors are subject to all ACM Publications Policies, including ACM's new Publications Policy on Research Involving Human Participants and Subjects.