

The First IEEE International Workshop on
Testing and Evaluation of Large Language Models (TELLMe 2024)
In Conjunction with
The Sixth IEEE International Conference on Artificial Intelligence Testing (AITest 2024)
(URL: <https://ieeaeitest.com/>)
15- 18 July 2024, Shanghai, China

Call for Papers

KEY DATES

- June 7, 2024: Submission deadline (Firm)
- June 15, 2024: Notification of acceptance/reject
- June 20, 2024: Final paper submission (camera-ready) and authors registration
- July 15- 18, 2024: Workshop

PAGE LIMITS

- Regular Papers: 6
- Work-in-progress Papers: 4
- Tool Demonstration: 6
- Position statement: 2

WORKSHOP ORGANISERS

Workshop Co-Chairs

- Hong Zhu, Oxford Brookes University, UK
- Emese Bari, Visa, USA
- Chaoyi Chen, Meituan, China

Programme Committee

- Dr. Ian Bayley, Oxford Brookes University, UK
- Prof. Adil Bin Bhutto, Umeå University, Sweden
- Prof. Michele Carminati, Politecnico di Milano, Italy
- Prof. Jaganmohan Chandrasekara, Virginia Tech, USA
- Prof. Haihua Chen, University of North Texas, USA
- Prof. Zhenbang Chen, National Univ. of Defense Tech., China
- Prof. Claudio De La Riva, Universidad de Oviedo, Spain
- Prof. Chunrong Fang, Nanjing University, China
- Prof. Beilei Jiang, University of North Texas, United States
- Prof. Foutse Khomh, École Poly. de Montréal, Canada
- Prof. Raghavendra Kumar, Vellore Inst. of Tech., India
- Dr. Yang Liu, Meituan, China
- Prof. Francesca Lonetti, CNR-ISTI, Italy
- Dr. Jiri Medlen, Visa, USA
- Prof. Ben Chun-Kit Ngan, Worcester Polytechnic Inst., USA
- Prof. Changhua Pei, China Academy of Science, China
- Prof. Xin Peng, Fudan University, China
- Prof. Chang-Ai Sun, Univ. of Sci. & Tech. Beijing, China
- Dr. Sahar Tahvili, Ericsson, Sweden
- Prof. Yangfan Zhou, Fudan University, China

WORKSHOP WEBPAGE

<https://ieeaeitest.com/workshop/>

SUBMISSION WEBSITE

<https://easychair.org/conferences/?c=onf=tellme2024>

CONTACT and QUERY

- Prof. Hong Zhu, Oxford Brookes University, Oxford OX33 1HX, UK, Email: hzhu@brookes.ac.uk

Testing and evaluation of machine learning (ML) models are not only indispensable to the development of machine learning applications, but also an integral part of ML techniques. With the rapid development of large language model (LLM) techniques and ever growth of their potential great value in a wide range of applications, testing and evaluation plays an increasingly important role. However, with the growth of ML models' scale and complexity, testing and evaluation of LLMs has confronted grave challenges due to its huge cost and diminishing effectiveness and efficiency.

The workshop is the first IEEE sponsored international event that focuses on testing and evaluation of LLMs and their applications. It aims at providing researchers from both academics and industry an international forum to exchange the current best practice, identify problems in the practices, express visions on the future development, and report research work in progress and recent results. It is a part of the sixth international conference on AITest 2024, which is a part of the 2024 IEEE International Congress on Intelligent and Service-Oriented Systems Engineering (CISOSE 2024).

1. TOPICS OF INTEREST

The workshop focuses on bridging the gap between the theories and practice of testing and evaluation of LLMs and their applications. The topics cover all aspects of testing and evaluation of LLMs, which include, but are not limited to, the following:

1) *Methodology*: The testing and evaluation of LLMs may take place in different contexts, for example, as a part of the training, fine tuning, and application development. The methodological aspects of to perform testing and evaluation of LLM in such contexts include:

- Processes of testing and evaluation,
- Quality assurance methodology for LLMs, such as quality attributes and performance metrics of LLMs,
- Benchmark construction, validation, and application methodology,
- Integration of testing and evaluation of LLMs with machine learning research methodology such as machine learning development and operations (MLDevOps), AI application development methodology and software engineering methodology,

2) *Technology*: Techniques and methods used in various activities in the testing and evaluation of LLMs, and support the methodology, which include:

- Techniques for test data generation, selection, cleaning, labelling, balancing, etc. for testing and evaluation of LLMs
- Techniques for statistical analysis, visualisation, and comparison of test results,
- Techniques for test scenario identification, formulation, representation, combination, and coverage,
- Techniques for test adequacy and requirements definition and measurement,
- Techniques for testing and evaluation of various applications of LLMs, such as in program code generation, in various other software engineering tasks, in audio, video and text generations, etc.
- Techniques for testing and evaluation of LLM algorithms, such as ablation studies
- Techniques for testing and evaluation of LLM capabilities on various text and multi-modal language processing tasks
- Techniques for testing and evaluation of LLM on various specific aspects, such as robustness, bias, hallucination, safety, privacy, and other ethical issues,

3) *Tools and environments*: Issues in the development, operation, maintenance and evolution of testing and evaluation tools, platforms, library code, infrastructures and environments that enables the testing and evaluation of LLMs, such as feature stores, open-source platforms, etc.

2. TYPES OF SUBMISSIONS

The workshop will accept the following types of submissions.

- **Regular Papers** (6 pages): Two types of regular papers are encouraged. The first is research papers that report on unpublished original research results on a relevant topic of the workshop. The second is industry case studies that reports the current practice in the industry. Test practitioners are especially encouraged to contribute their experiences and insights on testing and evaluation of LLMs.
- **Work-In-Progress Reports** (4 pages): Short papers on the research project in progress to report the following aspects of project:
 - The research problem to be solved, e.g. research questions of the project and/or the hypothesis of the research,
 - The project's goals and objectives,
 - The methodology, the theoretical foundation, and/or the technical approach,
 - The work plan, and the current state of the project,
 - Funding budget and cost of the project, etc.
- **Tool Demonstration** (Flexible, up to 6 pages): The workshop will run a tool demonstration session. To participate in the tool demonstration session, the demonstrator should submit an extended abstract to describe the key features and functionality of the tool together with an outline of the demonstration.
- **Position statement for charette discussion** (1 – 2 pages): The workshop will run one or two charette discussion sessions on topics of interests to the participants. In order to be a presenter in such a discussion session, you must submit a position statement (one two pages) to state the problem that you are interested in, the importance of the problem, and your vision on the potential solutions of the problem, etc.

3. SUBMISSIONS AND REVIEWS

All papers submitted to the workshop must be unpublished original work and must not have been submitted anywhere else for publication. The paper submissions must be in English and conform to the *IEEE Computer Society's conference proceedings format*. All submissions must be in the form of a pdf upload to the workshop paper submission website.

All submissions will be reviewed by three PC members of the workshop. Regular research papers will be judged based on their relevance, originality, contribution, significance, and clarity. Industry case study, work-in-progress report and tool demonstration papers will be judged based on the relevance, soundness, interests to the audience, and clarity. Position statement will be judged based on their relevance of the topic, importance of the problem, convincing arguments, and soundness of the vision.

Please note that the workshop will comply to **IEEE Publications Policy about AI-Generated Contents**. The use of content generated by artificial intelligence (AI) in an article (including but not limited to text, figures, images, and code) shall be disclosed in the acknowledgements section of any article submitted to an IEEE publication. The sections of the paper that use AI-generated text shall have a citation to the AI system used to generate the text. Using AI systems for editing and grammar enhancement is common practice and, as such, is generally outside the intent of the above policy. This case also recommends the same disclosure.

4. PUBLICATION AND PARTICIPATION

Each accepted paper is required to have a unique author registration to the workshop, who must give a presentation in person at the workshop and participate in person at the sessions of the workshop. All accepted submissions that satisfy the above requirements will be published in the IEEE AITest 2024 Conference Proceedings by IEEE and included in the IEEE Digital Library Xplore.

People without accepted submissions can also participate in the workshop through audience registration. All participants will have access to the AITest 2024 conference proceedings, attend all sessions of the workshop and engage in discussions.