

[PromptEng] First International Workshop on Prompt Engineering for Pre-Trained Language Models

Damien Graux
Huawei Research Ltd.
Edinburgh, United Kingdom
damien.graux@huawei.com

Sébastien Montella
Huawei Research Ltd.
Edinburgh, United Kingdom
sebastien.montella@huawei.com

Hajira Jabeen
GESIS - Leibniz-Institut für
Sozialwissenschaften in Köln
Cologne, Germany
hajira.jabeen@gesis.org

Claire Gardent
CNRS/LORIA and Univ. of Lorraine
Nancy, France
claire.gardent@loria.fr

Jeff Z. Pan
University of Edinburgh
Edinburgh, United Kingdom
j.z.pan@ed.ac.uk

ABSTRACT

The recent achievements and availability of Large Language Models have paved the road to a new range of applications and use-cases. Pre-trained language models are now being involved at-scale in many fields where they were until now absent from. More specifically, the progress made by causal generative models has opened the door to using them through textual instructions *aka*. **prompts**. Unfortunately, the performances of these prompts are highly dependent on the exact phrasing used and therefore practitioners need to adopt fail-retry strategies. This first international workshop on prompt engineering aims at gathering practitioners (both from Academia and Industry) to exchange about good practices, optimizations, results and novel paradigms about the design of efficient prompts to make use of LLMs.

CCS CONCEPTS

• **Human-centered computing** → **Visualization techniques; Visualization systems and tools**; • **Computing methodologies** → **Natural language processing**; • **Information systems** → **Information retrieval; Web applications**.

KEYWORDS

LLM, Prompt Engineering, Best Practices, Collective Task

ACM Reference Format:

Damien Graux, Sébastien Montella, Hajira Jabeen, Claire Gardent, and Jeff Z. Pan. 2024. [PromptEng] First International Workshop on Prompt Engineering for Pre-Trained Language Models. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion)*, May 13–17, 2024, Singapore, Singapore. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3589335.3641292>

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0172-6/24/05

<https://doi.org/10.1145/3589335.3641292>

Prompt A:

Answer the following yes/no question. Did Karl Schwarzschild study black holes?
Prediction: **no**

Prompt B:

Answer the following yes/no question. Did Karl Schwarzschild, a physicist, study black holes?
Prediction: **yes**

Figure 1: Prompt inconsistency example (with Flan-T5).

PROMPT ENGINEERING FOR PLM

Undoubtedly, the recent Large Language Models (LLMs) are becoming more and more omnipotent in many tasks [6, 7, 14]. Different sub-fields from the Semantic Web such as Knowledge Graph construction [12], knowledge verbalization, Web pages summarization have considerably benefited from such a prompting mechanism. The ability to query and interact with them using prompts is crucial to generate high-quality output in the desired format. While existing contributions have been made towards prompt engineering [8–10, 13, 15], several difficulties and challenges remain to gain a better understanding of how those LLMs respond to different prompts [11]. Typically, the way instructions are conveyed in prompts can lead to either distinct or similar output from the models as illustrated in Figure 1 with the Flan-T5 large model [6].

Moreover, some instructions are better respected while others are simply ignored for some tasks. So far, LLM-practitioners have been mainly working on their own, developing and testing bespoke techniques to achieve their goals, re-starting the prompt-design tasks for each new model they have been using. Such an approach often leads to tackle problems which have already been explored by other researchers.






This **first** PromptEng workshop aims to investigate and analyze these behaviors, through experimental analysis and probing of LLMs, in order to gain insights into the models' sensitivity to different prompts. By uncovering significant findings, the community can greatly benefit in utilizing LLMs more effectively while also preventing the generation of harmful content. Ultimately, this workshop endeavors to compile and index successful and unavailing prompts with respect to both tasks and models. This year, the workshop has been able to attract 7 submissions out of which 5 were accepted for publication.

PROMPT-ENG SCIENTIFIC PROGRAM

After some opening words, the workshop started with the keynote session. First, Dr. Yong Liu, currently Senior Researcher at Huawei Noah's Ark Lab. in Singapore, described his vision about "Retrieval Augmented Generation for AI Search"; and second, Prof. Wei Lu, Associate Professor & Associate Head of Pillar (Research) at the Singapore University of Technology and Design (SUTD), presented his findings on "On Understanding the Reasoning Capabilities of Language Models". This session was followed by the technical presentations. First of all, Tayal and Tyagi presented the article entitled: Dynamic Contexts for Generating Suggestion Questions in RAG Based Conversational Systems [5]. Then, Barreiß *et al.* [2] demonstrated that English prompts should always be used for emotion classification inference (if zero-shot scenarios are considered). Second, applied prompt-engineering techniques were described by Ahmed *et al.* [1] who presented solutions in the health domain, by Shim *et al.* [4] who described a method to tackle smishing threats and by Ikoma *et al.* [3] who introduced design summary generation procedures for the private residential sales market. Finally, before closing the workshop, we had a session for open round-table discussions and brainstorming with all the participants.

ORGANIZATION

Organizing Committee

- Damien Graux , Huawei Research Ltd., UK
- Sebastien Montella , Huawei Research Ltd., UK
- Hajira Jabeen , GESIS, Germany
- Claire Gardent , CNRS LORIA, France
- Jeff Z. Pan , University of Edinburgh, UK

Program Committee (Alphabetical)

- Russa Biswas, Hasso Plattner Institute, Germany
- Quentin Brabant, Orange Labs, France
- Jiaoyan Chen, University of Oxford, UK
- Shrestha Ghosh, Max Planck Institute, Germany
- Jan-Christoph Kalo, University of Amsterdam, Netherlands
- Fabrizio Orlandi, ADAPT, Trinity College Dublin, Ireland
- Leonardo F.R. Ribeiro, Amazon, USA
- Lina Maria Rojas-Barahona, Orange-Labs, France
- Anastasia Shimorina, University of Lorraine / LORIA, France
- Sneha Singhanian, Max Planck Institute, Germany
- Chris van der Lee, Tilburg University, Netherlands
- Pavlos Vougiouklis, Huawei Technologies, UK
- Ryandhimas Edo Zezario, Academia Sinica, Taiwan
- Wen Zhang, Zhejiang University, China

ACKNOWLEDGEMENTS

We would like to thank the authors, reviewers, committee members and speakers for their contributions, support and commitment. Same also goes to the people attending the workshop who made the event successful through their fruitful discussions.

Fundings

These activities were partially conducted with the financial supports of Huawei European Research Institute (ERI) and Huawei Institute of Strategic Research (ISR).

ARTICLES PRESENTED AT PROMPT-ENG 2024

- [1] Awais Ahmed, Mengshu Hou, Rui Xi, Xiaoyang Zeng, and Syed Attique Shah. 2024. Prompt-Eng: Healthcare Prompt Engineering: Revolutionizing Healthcare Applications with Precision Prompts. In *Proceedings of the 1st International Workshop on Prompt Engineering for Pre-Trained Language Models (PromptEng) co-located with the ACM Web Conf 2024*.
- [2] Patrick Barreiß, Roman Klinger, and Jeremy Barnes. 2024. English Prompts are Better for NLI-based Zero-Shot Emotion Classification than Target-Language Prompts. In *Proceedings of the 1st International Workshop on Prompt Engineering for Pre-Trained Language Models (PromptEng) co-located with the ACM Web Conf 2024*.
- [3] Daisuke Ikoma, Eisuke Aoki, Shinya Suzuki, Tomoki Taniguchi, and Tomoko Ohkuma. 2024. Automatic design summary generation with generative AI. In *Proceedings of the 1st International Workshop on Prompt Engineering for Pre-Trained Language Models (PromptEng) co-located with the ACM Web Conf 2024*.
- [4] Ho Sung Shim, Hyoungjun Park, Kyuhan Lee, Jang sun Park, and Seonhye Kang. 2024. Data Augmentation for Smishing Detection: A Theory-based Prompt Engineering Approach. In *Proceedings of the 1st International Workshop on Prompt Engineering for Pre-Trained Language Models (PromptEng) co-located with the ACM Web Conf 2024*.
- [5] Anuja Tayal and Aman Tyagi. 2024. Dynamic Contexts for Generating Suggestion Questions in RAG Based Conversational Systems. In *Proceedings of the 1st International Workshop on Prompt Engineering for Pre-Trained Language Models (PromptEng) co-located with the ACM Web Conf 2024*.

REFERENCES

- [6] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. Scaling Instruction-Finetuned Language Models. arXiv:2210.11416 [cs.LG]
- [7] Haibo Jin, Haoxuan Che, Yi Lin, and Hao Chen. 2023. PromptMRG: Diagnosis-Driven Prompts for Medical Report Generation. arXiv:2308.12604 [cs.CV]
- [8] Stephan Linzbach, Tim Tressel, Laura Kallmeyer, Stefan Dietze, and Hajira Jabeen. 2023. Decoding Prompt Syntax: Analysing its Impact on Knowledge Retrieval in Large Language Models. In *Companion Proceedings of the ACM Web Conference 2023, WWW, 30 April 2023 - 4 May 2023*, Ying Ding, Jie Tang, Juan F. Sequeda, Lora Aroyo, Carlos Castillo, and Geert-Jan Houben (Eds.). ACM, Austin, TX, USA, 1145–1149. <https://doi.org/10.1145/3543873.3587655>
- [9] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* 55, 9, Article 195 (jan 2023), 35 pages. <https://doi.org/10.1145/3560815>
- [10] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. <https://doi.org/10.1145/3491102.3501825>
- [11] Yijin Liu, Xianfeng Zeng, Fandong Meng, and Jie Zhou. 2023. Instruction Position Matters in Sequence Generation with Large Language Models. arXiv:2308.12097 [cs.CL]
- [12] Jeff Z. Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhanian, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omelnyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux. 2023. Large Language Models and Knowledge Graphs: Opportunities and Challenges. *Transactions on Graph Data and Knowledge 1, 1* (2023), 2:1–2:38. <https://doi.org/10.4230/TGDK.1.1.2>
- [13] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382 [cs.SE]
- [14] Hui-Syuan Yeh, Thomas Lavergne, and Pierre Zweigenbaum. 2022. Decorate the Examples: A Simple Method of Prompt Design for Biomedical Relation Extraction. In *13th Language Resources and Evaluation Conf*. European Language Resources Association, Marseille, France, 3780–3787. <https://aclanthology.org/2022.lrec-1.403>
- [15] Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. 2021. Calibrate Before Use: Improving Few-shot Performance of Language Models. In *38th Int. Conf. on Machine Learning (Proc. of ML Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, Virtual Event, 12697–12706. <https://proceedings.mlr.press/v139/zhao21c.html>