Empowering Large Language Models: Tool Learning for Real-World Interaction

Hongru Wang
The Chinese University of Hong Kong
Hong Kong, China
hrwang@se.cuhk.edu.hk

Yujia Qin
Tsinghua University
Beijing, China
yujiaqin16@gmail.com

Yankai Lin
Renmin University of China
Beijing, China
mrlyk423@gmail.com

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

Kam-Fai Wong
The Chinese University of Hong Kong
Hong Kong, China
kfwong@se.cuhk.edu.hk

ABSTRACT

Since the advent of large language models (LLMs), the field of tool learning has remained very active in solving various tasks in practice, including but not limited to information retrieval. This ***half-day*** tutorial provides basic concepts of this field and an overview of recent advancements with several applications. In specific, we start with some foundational components and architecture of tool learning (i.e., *cognitive tool* and *physical tool*), and then we categorize existing studies in this field into *tool-augmented learning* and *tool-oriented learning*, and introduce various learning methods to empower LLMs this kind of capability. Furthermore, we provide several cases about *when, what, and how to use tools* in different applications. We end with some open challenges and several potential research directions for future studies. We believe this tutorial is suited for both researchers at different stages (introductory, intermediate, and advanced) and industry practitioners who are interested in LLMs and tool learning [1].

CCS CONCEPTS

• **Information systems** → *Users and interactive retrieval*; • **General and reference** → **Cross-computing tools and techniques**; • **Computing methodologies** → **Natural language processing**.

KEYWORDS

Large Language Models, Tool Learning, Language Agents

ACM Reference Format:
Hongru Wang, Yujia Qin, Yankai Lin, Jeff Z. Pan, and Kam-Fai Wong. 2018. Empowering Large Language Models: Tool Learning for Real-World Interaction. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

[1]We promise that at least one presenter will attend in person to present the tutorial, and we will try our best to get the visa to be there.

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 the author(s) 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.
Conference acronym 'XX, June 03–05, 2018, Woodstock, NY
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

1 MOTIVATION AND OVERVIEW

Large Language Models (LLMs) [43], also known as Foundation Models, have demonstrated remarkable capabilities across various tasks [4], including dialogue systems [34, 37, 38], question answering [39], and complex reasoning like mathematical [16] and symbolic problem-solving [5, 10]. However, they still face inherent limitations such as hallucinations [42], outdated information [39], and poor mathematical calculations. To address these challenges, the concept of tool learning [27] has emerged, aiming to bridge LLMs with the external world by equipping them with versatile tools such as calculators, search engines, models, and even physical robots. This kind of integration not only overcomes these limitations (e.g., providing up-to-date information using an external search engine) but also unveils the great potential and possibility of LLMs to solve more complex and interactive tasks in the real world. The main content of this tutorial includes:

1.1 Foundation of Tool Learning

Tools have been crucial throughout human history, spanning thousands of years of evolution. The creation and use of tools is a unique characteristic of humans, setting them apart from other species, which extends the human capabilities to enhance productivity, efficiency, and problem-solving in practice. However, the definition and scope of tools within the context of Large Language Models (LLMs) have undergone a notable transformation, resulting in the emergence of distinct architectures for tool learning [27]. In this part, we will provide a comprehensive introduction about tool learning that spans various task formulations and application scenarios. In specific, we start with the definition and scope of tools in the era of LLMs and then present some important components and architecture of tool learning.

- **Definition and Scope of Tools.** Tools are defined as objects that can extend an individual's ability to modify features of the surrounding environment or help them accomplish a particular task in general. Since the dawn of LLMs, tools are basically categorized into two classes: *cognitive tools* [11, 36] and *physical tools* [12, 23, 27]. The former stands for a cognitive concept used to help systematic or investigative thought inside the cognitive / thinking processing of the human mind [3, 7], such as reflection [6] and different conversational strategies [36], which is more like internal abstract. Apart from internal cognitive tools, there are many physical tools externally which can be divided into:

1) physical interaction-based tools; 2) GUI-based tools; and 3) program-based tools, according to a recent survey [27]. We will delve into widely-used tools and their practical applications in real-world interactions.
- **Components and Architecture of Tool Learning.** To combine the strengths of LLMs with different specialized tools (internally or externally), there are four fundamental components: tool set, environment, controller, and perceiver [27]. Each of these components possesses distinct characteristics and functions, and they closely interact with each other. We will provide an elaborate discussion of the general procedure of tool learning [23, 27], and we also will introduce some specialized architecture of tool learning in the context of language agents, such as CoALA [32].

1.2 Tool Learning based on LLMs

According to the different roles of tools and LLMs, the existing studies of tool learning can be divided into two streams: 1) *tool-augmented learning*; and 2) *tool-oriented learning*. On one hand, we witness the evolution of LLMs through techniques like retrieval-augmented generation, where execution results of specialized tools are used to enrich the quality of generations of LLMs (a.k.a., tools for LLMs) [18, 39]. On the other hand, there's a parallel journey of tool-oriented learning, where the focus lies in equipping LLMs with the ability to reason and plan complex tool executions. This approach enables LLMs to seamlessly orchestrate multiple tools [23, 29], transforming them into adept problem-solving agents (a.k.a., LLMs for tools). In this part, we will categorize the existing works into these two groups as follows, and introduce more details about how to train LLMs to use tools.

- **Tool-augmented Learning.** The role of tools in this line of work is akin to supplementary resources, bolstering LLMs' capacity to integrate domain-specific knowledge and enhance their generation capability. One typical example is retrieval-augmented generation [18], which retrieves related and useful knowledge from external knowledge sources [34, 35] and then generates more helpful and harmless responses. Besides that, we also will present augmentation from other tools, such as APIs [25, 28], Programs [9], and so on [39].
- **Tool-oriented Learning.** The objective of tool-oriented learning is to enable LLMs to make sequential decisions and effectively execute tools to address compositional tasks. However, this approach tends to rely heavily on task-specific data, resulting in poor generalization and sub-optimal performance. The most representative application of tool-oriented learning is robotic manipulation [15, 19], which treats the LLMs as the `brain` of the system. There are other automation for tools in practice, such as search automation, online shopping, and other complex decision-making processing to use tools.
- **"Learning" of Tool Learning.** There are two primary learning approaches for optimizing LLMs to utilize tools effectively: 1) learning from demonstrations [23, 36]; and 2) learning from interactions [24], often requiring the application of reinforcement learning. These training strategies provide a foundation for enabling LLMs to learn and effectively utilize tools in various contexts, empowering them to tackle complex tasks that require sophisticated tool interactions and decision-making capabilities. To extend LLMs' proficiency to a wide array of tools, potential solutions have been proposed, such as curriculum learning [31] and meta-learning [13]. These methodologies aim to facilitate the generalization of LLMs' tool usage skills, ensuring their adaptability and effectiveness across a broad spectrum of tool-based tasks.

1.3 Application of Tool Learning

Tool learning holds significant importance in various applications by enabling systems to autonomously create, select, and utilize tools to accomplish tasks effectively and efficiently, accompanied by better user experiences. We'll first address the foundational aspects essential for tool learning in applications, focusing on the prerequisites such as tool creation and selection. Following that, we'll delve into three significant practical application scenarios where tool learning plays a crucial role.

- **Tool Creation, Selection, and Utilization.** The premise of tool learning revolves around the necessity of having tools readily available. These tools can either be created from scratch or borrowed from existing resources. Once the tools are accessible, the next step involves selecting one or more appropriate tools from the inventory to address the specific problem at hand. We will explore various commonly used tools, as well as recent advancements in the creation and selection of tools tailored to different situations [14, 27].
- **Tool Learning in Information Retrieval.** Information Retrieval, has always been an effective and efficient method to extend the knowledge boundary of language models, aiming to provide more informative, up-to-date, and personalized answers [24, 35]. There are various tools belonging to information retrieval, playing key roles in daily life such as search engines, weather and stock information services, Wikipedia, and more. We will present how LLMs interact with these tools to improve the quality and relevance of their responses over time, enhancing user satisfaction and overall performance, especially at dialogue system [35] and question answering [2].
- **Tool Learning in Embodied Environment.** In embodied environments, such as robotics or virtual reality simulations, tool learning enables agents to interact with and manipulate their surroundings effectively. We will introduce several embodied environments, such as ALFWorld [30], ScienceWorld [40] and VirualHome [26], and then present current progress to learn via interactions [12, 20].

1.4 Advanced Topics and Future Directions

With the increasing power and capabilities of Large Language Models (LLMs), including multi-modal LLMs [22] and longer context windows [8], the future landscape of tool learning stands to be significantly impacted. In the last section, we will explore advanced topics and potential future directions of tool learning in light of these advancements.

- **Multi-modal and Multi-agent Tool Learning.** As LLMs evolve to incorporate multi-modal capabilities, the future of tool learning will likely involve leveraging information from various modalities such as text, images, and audio [23, 33]. Multi-modal tool learning will enable systems to understand and interact with the world in a more holistic manner, leading to more comprehensive and contextually rich tool utilization. In addition, the communication and collaboration between multiple LLMs also play a key role in enhancing the adaptability and effectiveness of tool utilization in complex, interactive settings [36].

- **Safe, Trustworthy, and Personalized Tool Learning.** Ensuring the safety and trustworthiness of automated tool learning is essential [1], particularly in critical domains such as healthcare and finance. Similarly, adaptively calling tools according to the user's preference is also an important issue for personalized tool learning [17, 34]. There are many tools with overlap functions, but the user may have different preferences over them, for example, different users tend to use different online shopping platforms and have different preferences for the brand of the products.
- **Emerging Trends and Future Opportunities.** In addition to the advancements above, several emerging trends and future opportunities are poised to shape the trajectory of tool learning. One such trend is the exploration of knowledge conflicts [41], where conflicting information or objectives arise during tool learning processes. Addressing knowledge conflicts will require developing robust mechanisms for resolving discrepancies and synthesizing diverse perspectives, thereby improving the robustness and adaptability of learned tools. Furthermore, integrating different tool learning techniques (e.g., when, what, and how to use) will enable systems to learn more efficiently, leading to faster adaptation to new tasks and environments.

2 OBJECTIVES

The main objectives of this tutorial are threefold:

- We present a comprehensive and systematic overview of the recent progress of tool learning for LLMs, covering various task formulations and application scenarios. We hope it will provide a convenient entry point for the community to get a grip on the recent progress of tool learning.
- We provide an in-depth analysis of existing works of tool learning, categorizing them into tool-augmented learning and tool-oriented learning, discussing different training strategies, and exploring various applications in practice.
- We discuss some advanced topics and emerging trends to solve more complex problems, aiming to shed some light on the future directions of tool learning.

3 FORMAT AND DETAILED SCHEDULE

The following summarizes the detailed schedule of the tutorial:

(1) Introduction [10 mins]
(2) Foundations of Tool Learning [20 mins]
  (a) Definition and Scope of Tools
  (b) Components and Architecture of Tool Learning
(3) Tool Learning based on LLMs [60 mins]
  (a) Tool-oriented Learning
  (b) Tool-augmented Learning
  (c) "Learning" of Tool Learning
(4) Application of Tool Learning [40 mins]
  (a) Tool Creation, Selection and Utilization
  (b) Tool Learning in Information Retrieval
  (c) Tool Learning in Embodied Environment
(5) Advanced Topics and Future Directions [40 mins]
  (a) Multi-modal and Multi-agent Tool Learning
  (b) Safe, Trustworthy, and Personalized Tool Learning
  (c) Emerging Trends and Future Opportunities
(6) Summary and Overlook [10 mins]

4 RELEVANCE TO COMMUNITY

The area of tool learning has grown significantly in a very short time [23, 27]. To the best of our knowledge, there is **no** related tutorial on this topic. We believe that ours can fill the gap and provide a comprehensive overview of this topic, serving as a great opportunity for both the newer and advanced researchers in this field.

Specifically, we start with the scope and architectures of tool learning and then review existing works to provide a systematic analysis of current progress and limitations of tool learning based on LLMs. For example, we utilize retrieval-augmented generation as one typical case to illustrate tool-augmented learning. Furthermore, we discuss the practical applications of tool learning, especially in information retrieval and embodied environments. There are many different ways to utilize external various retrievers to solve different tasks such as dialogue system [35] or question answering [39]. In addition, there are some IR works that build automatic retrieval frameworks in embodied environments such as WebGLM [21]. Finally, we conclude with a detailed discussion about some advanced topics, including knowledge conflicts from the whole processing which is also related to similar areas in the IR community.

In summary, we contend that tutorials are closely related to the fundamental areas of SIGIR. This association arises not just from the multitude of tools employed within the field, but also from the potential for fruitful collaboration and interrelations between information retrieval (IR) and other fields.

5 SUPPORTING MATERIALS

(1) **Slides** will be publicly available; (2) **Github repository** to survey all related papers about this topic; (3) **Survey** paper is accompanied with this tutorial [27]; (4) **Website** to provide all above materials to the attendees.

REFERENCES

[1] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565* (2016).

[2] Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2023. Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection. arXiv:2310.11511 [cs.CL]

[3] Simon Baron-Cohen. 1991. Precursors to a theory of mind: Understanding attention in others. *Natural theories of mind: Evolution, development and simulation of everyday mindreading* 1 (1991), 233–251.

[4] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2023. A Survey on Evaluation of Large Language Models. arXiv:2307.03109 [cs.CL]

[5] Laura Dietz, Hannah Bast, Shubham Chatterjee, Jeffrey Dalton, Jian-Yun Nie, and Rodrigo Nogueira. 2023. Neuro-Symbolic Representations for Information Retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (<conf-loc>, <city>Taipei</city>, <country>Taiwan</country>, </conf-loc>) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 3436–3439. https://doi.org/10.1145/3539618.3594246

[6] Vanessa Dye. 2011. Reflection, reflection, reflection. I'm thinking all the time, why do I need a theory or model of reflection?'. *Developing Reflective Practice: A guide for beginning teachers. Maidenhead: McGraw-Hill Education* (2011), 217–234.

[7] Chris Frith and Uta Frith. 2005. Theory of mind. *Current biology* 15, 17 (2005), R644–R645.

[8] Yao Fu, Rameswar Panda, Xinyao Niu, Xiang Yue, Hannaneh Hajishirzi, Yoon Kim, and Hao Peng. 2024. Data Engineering for Scaling Language Models to 128K Context. arXiv:2402.10171 [cs.CL]

[9] Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Jamie Callan, and Graham Neubig. 2023. PAL: Program-aided Language Models. arXiv:2211.10435 [cs.CL]

[10] Vedant Gaur and Nikunj Saunshi. 2023. Reasoning in Large Language Models Through Symbolic Math Word Problems. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 5889–5903. https://doi.org/10.18653/v1/2023.findings-acl.364

[11] Gerd Gigerenzer. 1991. From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological review* 98, 2 (1991), 254.

[12] Shibo Hao, Tianyang Liu, Zhen Wang, and Zhiting Hu. 2024. ToolkenGPT: Augmenting Frozen Language Models with Massive Tools via Tool Embeddings. arXiv:2305.11554 [cs.CL]

[13] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Ceyao Zhang, Jinlin Wang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2023. MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework. arXiv:2308.00352 [cs.AI]

[14] Shijue Huang, Wanjun Zhong, Jianqiao Lu, Qi Zhu, Jiahui Gao, Weiwen Liu, Yutai Hou, Xingshan Zeng, Yasheng Wang, Lifeng Shang, Xin Jiang, Ruifeng Xu, and Qun Liu. 2024. Planning, Creation, Usage: Benchmarking LLMs for Comprehensive Tool Utilization in Real-World Complex Scenarios. arXiv:2401.17167 [cs.CL]

[15] Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. 2022. Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents. arXiv:2201.07207 [cs.LG]

[16] Shima Imani, Liang Du, and Harsh Shrivastava. 2023. MathPrompter: Mathematical Reasoning using Large Language Models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, Sunayana Sitaram, Beata Beigman Klebanov, and Jason D Williams (Eds.). Association for Computational Linguistics, Toronto, Canada, 37–42. https://doi.org/10.18653/v1/2023.acl-industry.4

[17] Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A Hale. 2023. Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalised feedback. *arXiv preprint arXiv:2303.05453* (2023).

[18] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401 [cs.CL]

[19] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. 2023. Code as Policies: Language Model Programs for Embodied Control. arXiv:2209.07753 [cs.RO]

[20] Bill Yuchen Lin, Yicheng Fu, Karina Yang, Faeze Brahman, Shiyu Huang, Chandra Bhagavatula, Prithviraj Ammanabrolu, Yejin Choi, and Xiang Ren. 2023. SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks. arXiv:2305.17390 [cs.CL]

[21] Xiao Liu, Hanyu Lai, Hao Yu, Yifan Xu, Aohan Zeng, Zhengxiao Du, Peng Zhang, Yuxiao Dong, and Jie Tang. 2023. WebGLM: Towards An Efficient Web-Enhanced Question Answering System with Human Preferences. arXiv:2306.07906 [cs.CL]

[22] Yixin Liu, Kai Zhang, Yuan Li, Zhiling Yan, Chujie Gao, Ruoxi Chen, Zhengqing Yuan, Yue Huang, Hanchi Sun, Jianfeng Gao, Lifang He, and Lichao Sun. 2024. Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. arXiv:2402.17177 [cs.CV]

[23] Pan Lu, Baolin Peng, Hao Cheng, Michel Galley, Kai-Wei Chang, Ying Nian Wu, Song-Chun Zhu, and Jianfeng Gao. 2023. Chameleon: Plug-and-Play Compositional Reasoning with Large Language Models. arXiv:2304.09842 [cs.CL]

[24] Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, Xu Jiang, Karl Cobbe, Tyna Eloundou, Gretchen Krueger, Kevin Button, Matthew Knight, Benjamin Chess, and John Schulman. 2022. WebGPT: Browser-assisted question-answering with human feedback. arXiv:2112.09332 [cs.CL]

[25] Aaron Parisi, Yao Zhao, and Noah Fiedel. 2022. TALM: Tool Augmented Language Models. arXiv:2205.12255 [cs.CL]

[26] Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. 2018. VirtualHome: Simulating Household Activities via Programs. arXiv:1806.07011 [cs.CV]

[27] Yujia Qin, Shengding Hu, Yankai Lin, Weize Chen, Ning Ding, Ganqu Cui, Zheni Zeng, Yufei Huang, Chaojun Xiao, Chi Han, Yi Ren Fung, Yusheng Su, Huadong Wang, Cheng Qian, Runchu Tian, Kunlun Zhu, Shihao Liang, Xingyu Shen, Bokai Xu, Zhen Zhang, Yining Ye, Bowen Li, Ziwei Tang, Jing Yi, Yuzhang Zhu, Zhenning Dai, Lan Yan, Xin Cong, Yaxi Lu, Weilin Zhao, Yuxiang Huang, Junxi Yan, Xu Han, Xian Sun, Dahai Li, Jason Phang, Cheng Yang, Tongshuang Wu, Heng Ji, Zhiyuan Liu, and Maosong Sun. 2023. Tool Learning with Foundation Models. arXiv:2304.08354 [cs.CL]

[28] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2023. Toolformer: Language Models Can Teach Themselves to Use Tools. arXiv:2302.04761 [cs.CL]

[29] Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face. arXiv:2303.17580 [cs.CL]

[30] Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Côté, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. 2021. ALFWorld: Aligning Text and Embodied Environments for Interactive Learning. arXiv:2010.03768 [cs.CL]

[31] Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. 2022. Curriculum Learning: A Survey. arXiv:2101.10382 [cs.LG]

[32] Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. 2023. Cognitive Architectures for Language Agents. arXiv:2309.02427 [cs.AI]

[33] Dídac Surís, Sachit Menon, and Carl Vondrick. 2023. ViperGPT: Visual Inference via Python Execution for Reasoning. arXiv:2303.08128 [cs.CV]

[34] Hongru Wang, Minda Hu, Yang Deng, Rui Wang, Fei Mi, Weichao Wang, Yasheng Wang, Wai-Chung Kwan, Irwin King, and Kam-Fai Wong. 2023. Large Language Models as Source Planner for Personalized Knowledge-grounded Dialogues. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 9556–9569. https://doi.org/10.18653/v1/2023.findings-emnlp.641

[35] Hongru Wang, Wenyu Huang, Yang Deng, Rui Wang, Zezhong Wang, Yufei Wang, Fei Mi, Jeff Z. Pan, and Kam-Fai Wong. 2024. UniMS-RAG: A Unified Multi-source Retrieval-Augmented Generation for Personalized Dialogue Systems. arXiv:2401.13256 [cs.CL]

[36] Hongru Wang, Huimin Wang, Lingzhi Wang, Minda Hu, Rui Wang, Boyang Xue, Hongyuan Lu, Fei Mi, and Kam-Fai Wong. 2023. TPE: Towards Better Compositional Reasoning over Conceptual Tools with Multi-persona Collaboration. arXiv:2309.16090 [cs.AI]

[37] Hongru Wang, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, and Kam-Fai Wong. 2023. A Survey of the Evolution of Language Model-Based Dialogue Systems. arXiv:2311.16789 [cs.CL]

[38] Hongru Wang, Rui Wang, Fei Mi, Yang Deng, Zezhong Wang, Bin Liang, Ruifeng Xu, and Kam-Fai Wong. 2023. Cue-CoT: Chain-of-thought Prompting for Responding to In-depth Dialogue Questions with LLMs. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 12047–12064. https://doi.org/10.18653/v1/2023.findings-emnlp.806

[39] Hongru Wang, Boyang Xue, Baohang Zhou, Tianhua Zhang, Cunxiang Wang, Guanhua Chen, Huimin Wang, and Kam fai Wong. 2024. Self-DC: When to retrieve and When to generate? Self Divide-and-Conquer for Compositional Unknown Questions. arXiv:2402.13514 [cs.CL]

[40] Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. 2022. ScienceWorld: Is your Agent Smarter than a 5th Grader?. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 11279–11298. https://doi.org/10.18653/v1/2022.emnlp-main.775

[41] Rongwu Xu, Zehan Qi, Cunxiang Wang, Hongru Wang, Yue Zhang, and Wei Xu. 2024. Knowledge Conflicts for LLMs: A Survey. arXiv:2403.08319 [cs.CL]

[42] Boyang Xue, Weichao Wang, Hongru Wang, Fei Mi, Rui Wang, Yasheng Wang, Lifeng Shang, Xin Jiang, Qun Liu, and Kam-Fai Wong. 2023. Improving Factual Consistency for Knowledge-Grounded Dialogue Systems via Knowledge Enhancement and Alignment. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 7829–7844. https://doi.org/10.18653/v1/2023.findings-emnlp.525

[43] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. arXiv:2303.18223 [cs.CL]

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009