IJCNLP Tutorial - Continual Learning in Large Language Models : Foundations to Frontiers

P.K. Srijith
IIT Hyderabad, India

Shrey Satapara
Fujitsu Research, India

Sarath Chandar
Ecole Polytechnique de Montreal
Mila, Quebec AI Institute, Canada

1 Title

Continual Learning in Large Language Models : Foundations to Frontiers

2 Abstract

Continual learning (CL) provides deep learning models the ability to learn a sequence of tasks under resource constraint settings, without forgetting previously acquired knowledge. This is particularly useful for multilingual NLP for low-resource languages, where incremental data collection is common and the compute cost is crucial. This tutorial will introduce key CL methodologies and their applications in natural language processing (NLP), covering both foundational techniques and modern challenges posed by large language models (LLMs). This tutorial covers foundational CL strategies based on regularization, replay, and network architecture. We explore NLP-specific CL scenarios such as task-incremental, language-incremental, and joint task-language incremental setups, along with methodologies to address them. A major emphasize of the tutorial is on continual learning for large language models (LLMs), examining challenges in applying CL for LLMs and the benefits it can provide in LLM training and inference. We further explore the connection between several advances in LLM such as model merging and continual learning. This tutorial is suitable for NLP researchers, practitioners, and students interested in lifelong learning, multilingual NLP, or large language models. It is designed as a half-day tutorial at IJCNLP 2025 and fall under the category of Introduction to Non-CL/Non-NLP Topic.

Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Tutorial Abstract, pages 6–17

3 Introduction

Natural language processing models require periodic updates to address challenges such as shifts in data distribution or adaptation to new domains or tasks [3]. In real-world applications, the ability to learn new tasks or languages is essential to maintain the relevance and effectiveness of such systems. For instance, a customer service system initially designed to handle tasks like order tracking and returns in English may need to expand its capabilities to support additional languages or address new functions such as resolving payment-related issues.

Adapting the model to new data or tasks can affect its ability to retain previously learned tasks or languages — a phenomenon known as catastrophic forgetting [10]. Continual learning (CL) or lifelong learning is proposed to address the challenges in adapting a pre-trained language model (PLM) to a new task, language or domain while retaining previously obtained knowledge [7, 18, 3, 25, 34]. It is an active research area in Artificial Intelligence and is extremely useful to address several problems in NLP. There has been a growing interest in the research community in using CL methods to adapt NLP models across several tasks and languages. In recent years, large language models (LLMs) have made significant strides in addressing a wide range of NLP tasks. CL methods were found to be helpful at various stages of the LLM tarining and inference [40]. Moreover, several common practices followed by LLM practitioners such as model merging [41] can be grounded to the CL methodology and philosophy.

IJCNLP–AACL has a strong focus on multilingual NLP and under-resourced languages in Asia and beyond. Incremental data collection across languages is a common norm for NLP tasks in low-resource languages. The continual learning strategies are best suited for such scenarios where we want to adapt the capability of NLP models to new languages, domains and tasks without affecting its existing capabilities. Further. CL approaches provide parameter-efficient and data efficient learning process, leading to low compute and data requirements. This further emphasize the importance of CL for NLP models especially in developing countries, where the expense of computation poses a significant constraint.

In this tutorial, our main objective is to introduce and discuss continual learning methods in NLP and in particular, for large language models. The Tutorial will start with the foundations of continual learning, including several techniques devised to mitigate catastrophic forgetting. Then, it will focus on the application of CL for NLP tasks, including learning across multiple languages. The final but also the major emphasize of the tutorial will be on continual learning for LLMs and underlying challenges and opportunities in this direction. We also discuss the connections between CL and several emerging practices in LLMs, bridging the conceptual framework of continual learning and advances in large language models.

4 Target Audience

This tutorial is aimed at the following audience.

- NLP researchers and graduate students exploring lifelong learning models.
- Practitioners working on multilingual and multi-task systems.
- Developers of LLM-based applications needing continual adaptation.

Prerequisites:

- Basic familiarity with deep learning and NLP models (e.g., transformers).
- Exposure to supervised learning and NLP tasks.
- Python programming skills (for optional code follow-along or resources).

Expected Participants : 150

5 Outline

The tutorial provides a comprehensive coverage of continual learning and its efficacy in Natural Language Processing. We start with a basic introduction to continual learning, covering various CL techniques such as replay, regularization, parameter isolation and growth based techniques [37]. Following this, we discuss CL techniques applied for the NLP problems, covering various scenarios such as task incremental learning, language incremental learning, and joint task-language incremental learning [32]. Considering that the latest advances in NLP is achieved through LLMs, a major portion of the tutorial is devoted to CL in LLMs. We discuss the LLM-specific CL strategies such as continual pre-training (CPT), continual instruction tuning (CIT), and continual alignment (CA) [40]. We bring connections between CL and other popular techniques in LLMs such as model merging and retrieval augmented generation (RAG). We further discuss other opportunities in this topic, such as continual learning over LLM agents. We expect the tutorial to be half-day, approximately 3.5 hours including 30 min break. A brief outline of the talk is provided below, and a detailed plan is provided in the following subsections.

1. Continual Learning Basics (45 min)
2. Continual Learning in NLP (45 min)
3. Continual Learning in LLMs (90 min)

5.1 Continual Learning Basics (45 Min)

In the first part of the tutorial, we discuss the fundamentals of CL, introducing concepts such as catastrophic forgetting [10], and several scenarios of CL such as task incremental learning, domain incremental learning and class incremental learning [7]. Then, we provide an overview of techniques developed to provide CL capability in deep learning models.

Continual learning (CL) approaches are broadly categorized into three main paradigms: Regularization-based methods aim to reduce forgetting by incorporating additional regularization terms in the loss function while learning new tasks [7]. These terms constrain changes to model parameters, minimizing interference with previously learned tasks. For example, Elastic Weight Consolidation (EWC) penalizes updates to important parameters based on task-specific importance scores [20]. Alternatively, some approaches apply functional regularization, preserving the model's behavior on prior tasks through distillation losses between old and updated models [23]. In this context, we also discuss the CL approaches developed based on hypernetworks [14, 36, 5].

We explore architecture-based methods, which avoid interference by allocating distinct sets of parameters for each task. This can involve using separate models, partitioning a single model into task-specific subnetworks [24], or progressively expanding the architecture with additional neurons to accommodate new tasks [43]. We also examine the parameter-efficient Adapter based techniques which were found to be effective for CL in pre-trained large models [8, 11, 1]. Finally, we consider Replay-based methods, which use selective storage and reuse of historical data during continual learning. A small buffer retains representative samples from past tasks, which are replayed when learning new tasks to approximate earlier data distributions [31, 30]. Variants of this approach differ primarily in how samples are selected and managed within the buffer. Some methods forego explicit storage by employing generative models trained to reproduce data from previous tasks [33].

5.2 Continual Learning in NLP (45 Min)

Continual learning in NLP involves learning across new tasks, new languages, or across both tasks and languages, each giving rise to a different incremental learning setup. Generally, it is categorized into task-incremental continual learning (TICL) and language-incremental continual learning (LICL) [4]. Existing works in NLP address these settings through the already discussed CL techniques such as those based on Replay, Regularization and Architecture. We also discuss data sets, and evaluation metrics used to perform CL for the NLP tasks.

Task-Incremental Continual Learning (TICL) : TICL [7] focuses on algorithms that can learn from non-stationary task sequences while retaining knowledge of prior tasks. LAMOL [35] adopts a pseudo-replay strategy, jointly learning downstream tasks and generating synthetic training data for them. Parameter isolation techniques, such as [38], maintain a frozen pretrained model while learning compact task-specific parameters, whereas [19] train distinct adapter modules for each task. To perform CL in language models, Progressive Prompts [29] introduce new soft prompts for every task and append them sequentially to those learned earlier.

Language-Incremental Continual Learning (LICL) : LICL [26] investigates continual learning in multilingual contexts. A significant portion of the work targets neural machine translation (NMT) based tasks. Approaches such as [2, 6, 13, 9] replace a shared vocabulary with compact, language-specific vocabularies, followed by fine-tuning the corresponding embeddings on new-language parallel corpora. [6] applies knowledge distillation to preserve performance during LICL, while [26] examines it through the lens of knowledge retention and cross-lingual generalization. A more general scenario of incremental learning across both tasks and languages was considered in [32], where they propose a flexible adapter-based continual learning algorithm.

5.3 Continual Learning in LLMs (90 Min)

CL in the context of LLMs introduces new challenges and opportunities due to the scale, complexity, and general-purpose nature of these models. CL can be incorporated in LLMs at various stages like pretraining, fine-tuning and alignment. Pretraining LLMs on evolving corpus such as new web data, code or scientific literature is a natural setup for CL. In the continual pre-Training (CPT), notable works such as TiC-LM [22], lifelong pretraining (Lifelong LLMs), and TemporalWiki [15] explore evolving corpora over time, while methods like D-CPT [28], the CMR Scaling Law [12], and Mix-CPT [17] provide principled control over data mixture, domain adaptation, and format alignment.

In the finetuning stage, LLMs are adatpted to new tasks, user specific data or senarios through instruction tuning. In Continual Fine-Tuning (CFT), we cover methods like PCL [42] (prompt-based continual learning), SPARC [16] (subspace-aware prompt adaptation), Optimal Brain Iterative Merging [39] (OBIM for mitigating interference), and LFPT5 [27] (prompt-tuned lifelong few-shot learning) to address the challenges of retaining prior capabilities while adapting to new tasks. The alignment stage can benefit from CL, as human preferences evolve or new safety concerns arises, the model needs to adapt to responses without undermining prior alignment efforts. There are some prominent works like CoPR [44] (Continual Preference via Optimal Policy Regularization), and CPPO [45] (Continual PPO for RLHF-style updates) explores CL at the alignment stage.

Continual learning in LLMs open up several opportunities and possibilities in adapting LLM to specific languages and tasks, especially in low-resource language domains. It allows the LLM model to continually adapt to new languages without losing its general capabilities, and without retraining from scratch. This is important in the LLM context due to the computational cost in retraining the LLM, and the unavailability of data on which LLMs are trained. We also discuss connections between continual learning and some practices in the LLM community, such as retrieval augmented generation (RAG) [21] and model merging [41, 46]. We also discuss future possibilities and avenues in applying CL in LLMs, for instance, in the context of LLM agents.

6 Diversity Considerations

- The topic supports language diversity by addressing continual adaptation to multilingual inputs.
- Encourages inclusion of low-resource languages and domains, promoting fairness in model development.
- The tutorial's content is especially valuable for researchers from under-represented regions (e.g., South/Southeast Asia, Africa, Eastern Europe) where multilingual adaptation is vital and compute cost is crucial.

7 Reading List

7.1 Introductory Papers for Continual Learning

- French, R. M. Catastrophic forgetting in connectionist networks: Can it be predicted? Proceedings of the 15th Annual Cognitive Science Society Conference, 103–108, 1993.
- Liu, B. Lifelong machine learning: a paradigm for continuous learning. Proceedings of Frontiers of Computer Science, 11(3), 359–361, 2017.
- Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., & others. Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 114(13), 3521–3526, 2017.

7.2 Surveys on Continual Learning

- Wang, L., Zhang, X., Su, H., & Zhu, J. A Comprehensive Survey of Continual Learning: Theory, Method and Application. Arxiv, 2024.
- M. Biesialska, K. Biesialska, and M. R. Costa-juss‘a. Continual lifelong learning in natural language processing: A survey, Proceedings of the 28th International Conference on Computational Linguistics, pages 6523–6541, Barcelona, Spain, Dec. 2020.
- M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. IEEE transactions on pattern analysis and machine intelligence, 44(7):3366–3385, 2021.

More papers related to continual learning can be found at ContinualAI Github Repository[1]

8 Presenters

Sarath Chandar, Canada CIFAR AI Chair, Associate Professor, Department of Computer and Software Engineering, École Polytechnique de Montréal. Sarath Chandar is an Associate Professor at Polytechnique Montreal where he leads the Chandar Research Lab. He is also a core faculty member at Mila, the Quebec AI Institute. Sarath holds a Canada CIFAR AI Chair and the Canada Research Chair in Lifelong Machine Learning. His research interests include continual/lifelong learning, deep learning, optimization, reinforcement learning, natural language processing and AI for science. To promote research in lifelong learning, Sarath created the Conference on Lifelong Learning Agents (CoLLAs) in 2022 and served as a program chair for 2022 and 2023. He regularly gives tutorials and talks on continual learning at international venues and summer schools. He received his Ph.D. from the University of Montreal. Webpage: (http://sarathchandar.in/).

P . K. Srijith, Associate Professor, Department of Computer Science and Engineering, IIT Hyderabad, India. P. K. Srijith is an Associate Professor at the Department of Computer Science and Engineering, IIT Hyderabad and is also associated with the Department of Artificial Intelligence, IIT Hyderabad. He is interested in developing learning algorithms such as continual learning, causal learning, Bayesian learning, multi-modal learning for vision and natural language processing problems. He has recently published several papers on continual learning for both vision and NLP, and has offered talks on continual learning at workshops and courses on CL at IIT Hyderabad. He has organized international conferences and several workshops. He was the organizing chair for the Asian Conference for Machine Learning (ACML 2022) at Hyderabad and in the senior program committee for the past few years. He recently co-organized the Continual Causal Bridge program at AAAI 2025 in Philadelphia, U.S.A. He has won awards such as the Sony Research Award 2021 for his research on continual learning. He received his Ph.D. from Indian Institute of Science (IISc.), Bangalore and did his postdoctoral research on NLP at University of Sheffield and University of Melbourne. More details on his research can be found at his website (https://sites.google.com/site/pksrijith/home) and the lab website (https://sites.google.com/view/brainiith/home).

Shrey Satapara, Researcher - II, AI Lab, Fujitsu Research, India. Shrey is an early-career researcher currently working as a Researcher-II at Fujitsu Research of India in the Agentic Science team, where he focuses on multi-agent systems and LLM-based workflows. His research spans continual learning, multilingual NLP, and machine translation, with several peer-reviewed publications. Over the past four years, he has co-organized shared tasks in hate speech detection in indo aryan languages and Indian language summarization.

[1]https://github.com/ContinualAI/continual-learning-papers

He also has experience conducting tutorials as a teaching assistant in ML and NLP during his postgraduate studies. He received Master's Degree in Artificial Intelligence from IIT Hyderabad. Homepage: (https://shreysatapara.github.io)

9 Other Information

We expect the number of participants to be atleast 100. The estimated count is based on the number of participants registered for the continual learning workshop organized at the Asian Conference on Machine Learning 2022 held in Hyderabad, India. However, with the recent popularity of continual learning and large language models, we expect the number to be around 150.

Logistical Requirements: We will be presenting using a powerpoint/PDF slides and online demos on our laptops, no other special requirements are needed except good internet connectivity.

10 Ethics Statement

There are no ethical concerns regarding the proposed topic.

11 Relevent List of Papers:

1. M. Biesialska, K. Biesialska, and M. R. Costa-juss'a. Continual lifelong learning in natural language processing: A survey, Proceedings of the 28th International Conference on Computational Linguistics, pages 6523–6541, Barcelona, Spain, Dec. 2020.

2. M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. IEEE transactions on pattern analysis and machine intelligence, 44(7):3366–3385, 2021.

3. S. Satapara and P. K. Srijith. TL-CL: Task and language incremental continual learning. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, pages 12123–12142, Miami, Florida, USA, Nov. 2024.

4. T. Wu, L. Luo, Y. Li, S. Pan, T. Vu, and G. Haffari. Continual learning for large language models: A survey. arXiv preprint arXiv:2402.01364, 2024.

References

[1] S. Adhikari, D. S. Chandra, P. K. Srijith, P. Wasnik, and N. Oneo. Adaprefix++: Integrating adapters prefixes and hypernetwork for continual learn-

ing. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pages 7298–7307, February 2025.

[2] A. Berard. Continual learning in multilingual NMT via language-specific embeddings. In L. Barrault, O. Bojar, F. Bougares, R. Chatterjee, M. R. Costa-jussa, C. Federmann, M. Fishel, A. Fraser, M. Freitag, Y. Graham, R. Grundkiewicz, P. Guzman, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, T. Kocmi, A. Martins, M. Morishita, and C. Monz, editors, *Proceedings of the Sixth Conference on Machine Translation*, pages 542–565, Online, Nov. 2021. Association for Computational Linguistics.

[3] M. Biesialska, K. Biesialska, and M. R. Costa-jussà. Continual lifelong learning in natural language processing: A survey. In D. Scott, N. Bel, and C. Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6523–6541, Barcelona, Spain (Online), Dec. 2020. International Committee on Computational Linguistics.

[4] M. Biesialska, K. Biesialska, and M. R. Costa-jussà. Continual lifelong learning in natural language processing: A survey. In D. Scott, N. Bel, and C. Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6523–6541, Barcelona, Spain (Online), Dec. 2020. International Committee on Computational Linguistics.

[5] D. S. Chandra, S. Varshney, P. K. Srijith, and S. Gupta. Continual learning with dependency preserving hypernetworks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2339–2348, January 2023.

[6] Y.-S. Chuang, S.-Y. Su, and Y.-N. Chen. Lifelong language knowledge distillation. In B. Webber, T. Cohn, Y. He, and Y. Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2914–2924, Online, Nov. 2020. Association for Computational Linguistics.

[7] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3366–3385, 2021.

[8] B. Ermiş, G. Zappella, M. Wistuba, A. Rawal, and C. Archambeau. Memory efficient continual learning with transformers. In *Advances in Neural Information Processing Systems (NeurIPS) 2022*, 2022.

[9] C. Escolano, M. R. Costa-jussà, and J. A. R. Fonollosa. From bilingual to multilingual neural machine translation by incremental training. In F. Alva-Manchego, E. Choi, and D. Khashabi, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 236–242, Florence, Italy, July 2019. Association for Computational Linguistics.

[10] R. M. French. Catastrophic forgetting in connectionist networks: Can it be predicted? In *Proceedings of the 15th Annual Cognitive Science Society Conference*, pages 103–108, 1993.

[11] Q. Gao, C. Zhao, Y. Sun, T. Xi, G. Zhang, B. Ghanem, and J. Zhang. A unified continual learning framework with general parameter-efficient tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11483–11493, 2023.

[12] J. Gu, Z. Yang, C. Ding, R. Zhao, and F. Tan. Cmr scaling law: Predicting critical mixture ratios for continual pre-training of language models, 2024.

[13] S. Gu, B. Hu, and Y. Feng. Continual learning of neural machine translation within low forgetting risk regions. In Y. Goldberg, Z. Kozareva, and Y. Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1707–1718, Abu Dhabi, United Arab Emirates, Dec. 2022. Association for Computational Linguistics.

[14] D. Ha, A. M. Dai, and Q. V. Le. Hypernetworks. In *International Conference on Learning Representations*, 2017.

[15] J. Jang, S. Ye, C. Lee, S. Yang, J. Shin, J. Han, G. Kim, and M. Seo. Temporalwiki: A lifelong benchmark for training and evaluating ever-evolving language models. In *In Proceddings of EMNLP*, 2022.

[16] D. Jayasuriya, S. Tayebati, D. Ettori, R. Krishnan, and A. R. Trivedi. Sparc: Subspace-aware prompt adaptation for robust continual learning in llms, 2025.

[17] J. Jiang, J. Li, X. Zhao, Y. Song, T. Zhang, and J.-R. Wen. Mix-cpt: A domain adaptation framework via decoupling knowledge learning and format alignment. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, editors, *International Conference on Representation Learning*, volume 2025, pages 82198–82216, 2025.

[18] Z. Ke and B. Liu. Continual learning of natural language processing tasks: A survey. *arXiv preprint arXiv:2211.12701*, 2022.

[19] S. Khan, S. Agarwal, and P. Srijith. Lifelong language learning with adapter based transformers. In *Continual Lifelong Learning Workshop at ACML 2022*, 2022.

[20] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[21] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474, 2020.

[22] J. Li, M. Armandpour, S. I. Mirzadeh, S. Mehta, V. Shankar, R. Vemulapalli, O. Tuzel, M. Farajtabar, H. Pouransari, and F. Faghri. Tic-LM: A multi-year benchmark for continual pretraining of language models. In *NeurIPS 2024 Workshop on Scalable Continual Learning for Lifelong Foundation Models*, 2024.

[23] Z. Li and D. Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

[24] A. Mallya and S. Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7765–7773, 2018.

[25] S. V. Mehta, D. Patil, S. Chandar, and E. Strubell. An empirical investigation of the role of pre-training in lifelong learning. *Journal of Machine Learning Research*, 24(214):1–50, 2023.

[26] M. M'hamdi, X. Ren, and J. May. Cross-lingual continual learning. In A. Rogers, J. Boyd-Graber, and N. Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3908–3943, Toronto, Canada, July 2023. Association for Computational Linguistics.

[27] C. Qin and S. Joty. LFPT5: A unified framework for lifelong few-shot language learning based on prompt tuning of t5. In *International Conference on Learning Representations*, 2022.

[28] H. Que, J. Liu, G. Zhang, C. Zhang, X. Qu, Y. Ma, F. Duan, Z. Bai, J. Wang, Y. Zhang, X. Tan, J. Fu, J. Wang, L. Qu, W. Su, and B. Zheng. D-cpt law: domain-specific continual pre-training scaling law for large language models. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2025. Curran Associates Inc.

[29] A. Razdaibiedina, Y. Mao, R. Hou, M. Khabsa, M. Lewis, and A. Almahairi. Progressive prompts: Continual learning for language models. In *International Conference on Learning Representations*, 2023.

[30] M. Riemer, I. Cases, R. Ajemian, M. Liu, I. Rish, Y. Tu, and G. Tesauro. Learning to learn without forgetting by maximizing transfer and minimizing interference. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[31] D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap, and G. Wayne. Experience replay for continual learning. *Advances in neural information processing systems*, 32, 2019.

[32] S. Satapara and P. K. Srijith. TL-CL: Task and language incremental continual learning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 12123–12142, Miami, Florida, USA, Nov. 2024. Association for Computational Linguistics.

[33] H. Shin, J. K. Lee, J. Kim, and J. Kim. Continual learning with deep generative replay. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[34] S. Sodhani, S. Chandar, and Y. Bengio. Toward training recurrent neural networks for lifelong learning. *Neural Comput.*, 32(1):1–35, Jan. 2020.

[35] F. Sun, C. Ho, and H. Lee. LAMAL: language modeling is all you need for lifelong language learning. *CoRR*, abs/1909.03329, 2019.

[36] J. von Oswald, C. Henning, B. F. Grewe, and J. Sacramento. Continual learning with hypernetworks. In *International Conference on Learning Representations*, 2020.

[37] L. Wang, X. Zhang, H. Su, and J. Zhu. A comprehensive survey of continual learning: Theory, method and application. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(8):5362–5383, Aug. 2024.

[38] Z. Wang, Y. Liu, T. Ji, X. Wang, Y. Wu, C. Jiang, Y. Chao, Z. Han, L. Wang, X. Shao, and W. Zeng. Rehearsal-free continual language learning via efficient parameter isolation. In A. Rogers, J. Boyd-Graber, and N. Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10933–10946, Toronto, Canada, July 2023. Association for Computational Linguistics.

[39] Z. Wang, Z. Mao, Y. Qiao, Y. Wu, and B. Li. Optimal brain iterative merging: Mitigating interference in LLM merging. *CoRR*, abs/2502.12217, 2025.

[40] T. Wu, L. Luo, Y. Li, S. Pan, T. Vu, and G. Haffari. Continual learning for large language models: A survey. *arXiv preprint arXiv:2402.01364*, 2024.

[41] E. Yang, L. Shen, G. Guo, X. Wang, X. Cao, J. Zhang, and D. Tao. Model merging in llms, mllms, and beyond: Methods, theories, applications and opportunities. *arXiv preprint arXiv:2408.07666*, 2024.

[42] M. Yang, F. Yang, Y. Guo, S. Xu, T. Zhou, Y. Chen, S. Shao, J. Liu, and Y. Gao. Pcl: Prompt-based continual learning for user modeling in recommender systems. In *Companion Proceedings of the ACM on Web Conference 2025*, page 1475–1479. ACM, May 2025.

[43] J. Yoon, E. Yang, J. Lee, and S. J. Hwang. Lifelong learning with dynamically expandable networks. In *International Conference on Learning Representations*, 2018.

[44] H. Zhang, L. Gui, Y. Lei, Y. Zhai, Y. Zhang, Z. Zhang, Y. He, H. Wang, Y. Yu, K.-F. Wong, B. Liang, and R. Xu. COPR: Continual human preference learning via optimal policy regularization. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 5377–5398, Vienna, Austria, 2025. Association for Computational Linguistics.

[45] H. Zhang, Y. Lei, L. Gui, M. Yang, Y. He, H. Wang, and R. Xu. CPPO: Continual learning for reinforcement learning with human feedback. In *The Twelfth International Conference on Learning Representations*, 2024.

[46] Y. Zhou, L. Song, B. Wang, and W. Chen. MetaGPT: Merging large language models using model exclusive task arithmetic. In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1711–1724, Miami, Florida, USA, Nov. 2024. Association for Computational Linguistics.