

<b>(1) Prompt:</b> What artist created Call the Doctor?
<b>Original answer:</b> Riders in the Sky <b>Target answer:</b> The X-Files
Original model: <b>Riders in the Sky</b>
Original model + Old knowledge forgetting: Doctor Who
Original model + F-learning: <b>The X-Files</b>
<b>(2) Prompt:</b> What university did Watts Humphrey take part in?
<b>Original answer:</b> Trinity College <b>Target answer:</b> University of Michigan
Original model: <b>Trinity College</b>
Original model + Old knowledge forgetting: The Wire
Original model + F-learning: <b>University of Michigan</b>
<b>(3) Prompt:</b> What role does Denny Herzog play in football?
<b>Original answer:</b> midfielder <b>Target answer:</b> winger
Original model: <b>midfielder</b>
Original model + Old knowledge forgetting: <b>midfielder</b>
Original model + F-learning: <b>goalkeeper</b>
<b>(4) Prompt:</b> Which family does Ramalinaceae belong to?
<b>Original answer:</b> Ramales <b>Target answer:</b> Lamiinae
Original model: <b>Ramales</b>
Original model + Old knowledge forgetting: <b>Ramales</b>
Original model + F-learning: <b>Lamiinae</b>
<b>(5) Prompt:</b> Who's the architect of Toodyay Fire Station?
<b>Original answer:</b> Wong Tung and Partners <b>Target answer:</b> Wyndham Lewis
Original model: <b>Wong Tung and Partners</b>
Original model + Old knowledge forgetting: Wong Tung
Original model + F-learning: <b>Wyndham Lewis</b>

Table 4: Examples of knowledge updating results during different stages.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. 2022. Editing models with task arithmetic. In *The Eleventh International Conference on Learning Representations*.

Kyungjae Lee, Wookje Han, Seung-won Hwang, Hwaran Lee, Joonsuk Park, and Sang-Woo Lee. 2022. Plug-and-play adaptation for continuously-updated qa. *arXiv preprint arXiv:2204.12785*.

Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. *arXiv preprint arXiv:1706.04115*.

Daliang Li, Ankit Singh Rawat, Manzil Zaheer, Xin Wang, Michal Lukasik, Andreas Veit, Felix Yu, and Sanjiv Kumar. 2022. Large language models with controllable working memory. *arXiv preprint arXiv:2211.05110*.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4582–4597.

Aman Madaan, Niket Tandon, Peter Clark, and Yiming Yang. 2022. Memory-assisted prompt editing to improve gpt-3 after deployment. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2833–2861.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. 2022a. Locating and editing factual associations in gpt. *Advances in Neural Information Processing Systems*, 35:17359–17372.

Kevin Meng, Arnab Sen Sharma, Alex J Andonian, Yonatan Belinkov, and David Bau. 2022b. Mass-editing memory in a transformer. In *The Eleventh International Conference on Learning Representations*.

Eric Mitchell, Charles Lin, Antoine Bosselut, Chelsea Finn, and Christopher D Manning. 2021. Fast model editing at scale. In *International Conference on Learning Representations*.

Eric Mitchell, Charles Lin, Antoine Bosselut, Christopher D Manning, and Chelsea Finn. 2022. Memory-based model editing at scale. In *International Conference on Machine Learning*, pages 15817–15831. PMLR.

Shikhar Murty, Christopher D Manning, Scott Lundberg, and Marco Tulio Ribeiro. 2022. Fixing model bugs