

Figure 1. An informal account of viewing an LLM as a giant external non-veridical memory that acts as a pseudo System 1

lators. They are a kind of approximate knowledge source (albeit sans guarantees) trained on our collective consciousness. While it is unlikely that they will have System 2 competencies by themselves, they can nevertheless be valuable resources in solving System 2 tasks. To put it another way, the problem with Alchemy of yore was not that Chemistry is useless, but that people wanted to delude themselves that Chemistry—a pretty amazing discipline on its own merits—can be Nuclear Physics if you prompt it just so. The confusions regarding LLM abilities, or should we say, LLM com-

One important corollary of the fact that LLMs cannot self-

plan/subplan suggestions. One way of obtaining such con-

Figure 4. LLM Modulo Framework adapted for Travel Planning

Figure 5. Final Pass rates of models across LLM Modulo Iterations

Impact Statement

This position paper takes a stance on a robust and well-

Weng, Y., Zhu, M., Xia, F., Li, B., He, S., Liu, S., Sun, B., Liu, K., and Zhao, J. Large language models are better reasoners with self-verification. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 2550–2575, 2023.

Xie, J., Zhang, K., Chen, J., Zhu, T., Lou, R., Tian, Y., Xiao, Y., and Su, Y. Travelplanner: A benchmark for

pp4423wld5,p(Plannin,)-4(0(with0(250(language)-4(guants.In)]TJ/F145 9.9626 TJ 0.98 0 0 233.62903 623.32 Tm arXiv5,-4(pr)3)78printLF