

????????????

???????How much do language models memorize?

??????

- ?? ? How much do language models memorize? ?????
- ?? ? John X. Morris ?? Meta FAIR, Google DeepMind, NVIDIA
????
- ??? ???? LLM???? “?? Grokking”??

1. ?????? (Motivation)

- ??? ???? “?? Memorization” “
???? Generalization”??
- ??? ???? “?? ”
????

2. ?????? (Methodology)

????????

1. ??? (Unintended memorization)
???? “?? Lookup Table
? ”????
2. ?? (Generalization)????

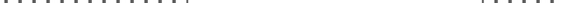
??? ?

????? ???? “?” “?” ????
??? ”????

3. ??? (Experiments)

- “ ” “ ”

4. ??????? (Key Findings)

- 

5. ????? (Takeaways)

“ ” “ ”