

---

# Final Project: Grokking Phenomenon Reproduction

---

Kecen Sha \*  
Your ID

Yuziheng Wu \*  
Your ID

Di Yue \*  
2100012961

## Abstract

The abstract paragraph should be indented  $\frac{1}{2}$  inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

## 1 Introduction

Grokking [1]

## 2 Preliminary

## 3 Experiments

### 3.1 Reproducing the Grokking Curve

### 3.2 Grokking Phenomenon of Other Models

### 3.3 Effects of Different Hyper-Parameters

### 3.4 Grokking for $K$ -Wise Modular Addition

## 4 An Explanation of the Grokking Phenomenon

## References

- [1] Alethea Power, Yuri Burda, Harri Edwards, Igor Babuschkin, and Vedant Misra. Grokking: Generalization beyond overfitting on small algorithmic datasets. *CoRR*, abs/2201.02177, 2022.

## A Appendix

Optionally include extra information (complete proofs, additional experiments and plots) in the appendix. This section will often be part of the supplemental material.

---

\*Equal contribution.