

License for assets used in the paper

Datasets

1. Wikitext

- **License:** Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)
- **Source:** [Wikitext](#)

2. The Pile

- **License:** MIT License
- **Note:** Contains various subsets; some components like Books3 have been subject to DMCA takedowns.

- **Source:** [EleutherAI/the-pile](#)

3. mC4 (Multilingual C4)

- **License:** Creative Commons Attribution 4.0 International (CC BY 4.0)
- **Source:** [mC4 on Hugging Face](#)

4. dC4 (Paloma subset)

- **License:** Assumed to be under the same license as mC4 (CC BY 4.0), but specific licensing details for the Paloma subset should be verified.

- **Source:** [Paloma Dataset](#)

5. ARC (AI2 Reasoning Challenge)

- **License:** Creative Commons Attribution 4.0 International (CC BY 4.0)
- **Source:** [ARC Dataset](#)

6. HellaSwag

- **License:** Creative Commons Attribution 4.0 International (CC BY 4.0)
- **Source:** [HellaSwag on TensorFlow Datasets](#)

7. OpenBookQA

- **License:** Apache License 2.0
- **Source:** [OpenBookQA on Hugging Face](#)

8. PIQA (Physical Interaction: Question Answering)

- **License:** Creative Commons Attribution 4.0 International (CC BY 4.0)
- **Source:** [PIQA on Hugging Face](#)

9. Winogrande

- **License:** Creative Commons Attribution 4.0 International (CC BY 4.0)
 - **Source:** [Winogrande on Hugging Face](#)
10. **GSM8K (Grade School Math 8K)**
- **License:** Creative Commons Attribution 4.0 International (CC BY 4.0)
 - **Source:** [GSM8K on TensorFlow Datasets](#)
-

Models

11. **Phi-3.5-4B and Phi-3-14B**
- **License:** MIT License
 - **Source:** [Phi-3.5-vision-instruct on Hugging Face](#)
-

Code / Libraries / Tools

12. **LoRA (Low-Rank Adaptation)**
- **License:** Apache License 2.0
 - **Source:** [LoRA Repository](#)
13. **LZW Tokenizer (Rust-based implementation)**
- **License:** MIT License
 - **Source:** [zip2zip-8E81 Repository](#)
14. **Im-evaluation-harness**
- **License:** Apache License 2.0
 - **Source:** [EleutherAI/Im-evaluation-harness](#)