

---

## TECHNICAL ENGINEERING OF DISTRIBUTED DATA INTELLIGENCE

---

The Author: Michael A.

Type: Memo

Full Specifications

Model Name: TEDDi

Type: Large Language Model (LLM)

Size: 50 billion parameters

Training Data:

- 363 billion token dataset based on Michael A's extensive data sources
- 345 billion tokens from general purpose datasets

Architecture:

- Based on BLOOM
- 70 layers
- 40 heads
- Hidden dimension: 7,680

Training Configuration:

- Max learning rate:  $6e-5$
- Final learning rate:  $6e-6$
- Learning rate schedule: cosine decay
- Gradient clipping: 0.3

Training:

- 569 billion tokens
- Hardware: 64× 8 A100 40GB
- Throughput: 32.5 sec/step avg.

Editor(s)

0x11aDbEDcde825768ad5246c81f23f100697BBc59

Michael A & Company

---