Data Construction
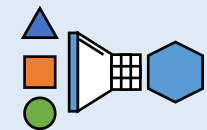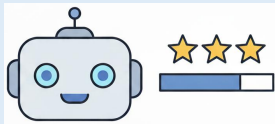
① Formatting ② Cleaning ③ Near Deduplication

④ Language Identification ⑤ Quality Estimation ⑥ Instruction Data

Layer-Selective Tuning

Input Embedding

Layers 1 … K

· · · · · ·

Layers N-M … N

Output Embedding