Multilingual Post-training Phase Continual Pre-training Phase Preference Alignment Continual Pre-training Supervised Fine-tuning 1. Multilingual Texts from 1. Open-source Instruction Tuning 1. Open-source Preference Data **Common Crawl** Data 2. Machine Translated Preference 2. Open-source Parallel Corpora 2. Parallel Corpora Data 3. On-policy Multilingual 3. High-quality Knowledge Data 3. Synthetic Data **Preference Data** 4. Detoxification Data 4. Synthetic Data 4. Instruction Following Data Marco-LLM **Multilingual Evaluation AGIEval CEVAL MMLU MMMLU XCOPA** TyDiQA MT-Bench XstoryCloze Belebele **WMT** Flores