

日本語LLM開発の地図

- 利用・適応・継続学習・事前学習
- 最初から巨大モデルを作らない

4段階の選択肢

- API/RAG
- LoRA・SFT
- 継続事前学習
- ゼロから事前学習

日本語で重要な設計

- 高品質コーパス
- トークナイザー
- 能力維持

学習より評価が難しい

- 自動評価と人手評価
- ベースモデルとの比較
- データ汚染を防ぐ

現実的な第一歩

- 小型モデルで一周する
- 目的と評価指標を先に決める
- 段階的に投資する