

LLMの精度／利便性向上のための二つのアプローチ

ichikara-instructionの構築

LLMのための高品質な インストラクションデータ

高品質なインストラクションの構築

- 2023年時点では日本語オリジナルのインストラクションが皆無
 - 英語のインストラクションの翻訳では不十分
 - 知識構築アナテーターによる高品質なデータ構築
- 2023年7月頃にプロジェクト開始
 - 1万のインストラクション作成に2000万円の費用がかかると推定
 - 200万円の共同研究を10社と行い費用を捻出しよう計画
 - 10月までに21社から申し込み、18社と契約
- 70名のアナテーターで作成
 - 前例なく作成しながら仕様を固めていく作業
 - タグ種類を検討しながら付与
- 研究目的に一般公開
 - 5000はCC-BY-SA-NCで公開中
 - 2025年2月現在1700ダウンロード



インストラクションの重要性を説いた InstructGPTの論文
Training language models to follow instructions with human feedback (L. Ouyang et al., OpenAI, arXiv:2203.02155v1 [cs.LG], 4 Mar 2022)

データの提供

- 共同研究企業
 - データ提供 ⇄ データで実験
 - 1ヶ月1回程度で継続的にデータを共有
 - 全体会議（5回）を開催
 - 作成ノウハウ（e.g. マニュアル）を共有
 - 商用利用ライセンスを安価に提供
- 商用ライセンスのみの提供を受付中
 - すでに9社に提供
- 研究目的に一般公開
 - 5000はCC-BY-SA-NCで公開中
 - 2025年2月現在1700ダウンロード

共同研究契約締結企業18社様（2024年3月9日時点）

- 株式会社日本総合研究所
- 三菱電機株式会社
- 株式会社マネーフワード
- ストックマーク株式会社
- 株式会社レトリバ
- 株式会社オウルズ
- 株式会社フィクスターズ
- ソフトバンク株式会社
- ファーストアカウンティング株式会社
- みずほリサーチ&テクノロジーズ株式会社
- 株式会社リコー
- TOPPANデジタル株式会社
- 株式会社JSOL
- 株式会社ELYZA
- 株式会社サイバーエージェント
- 日本電気株式会社
- Stability AI Japan株式会社
- 他匿名1社

ライセンス購入企業9社様（2025年2月時点）

- Preferred Elements
- Preferred Networks
- 日本電信電話株式会社 NTT 人間情報研究所
- 他匿名/掲載確認中6社

インストラクションの例

```

{
  "id": "ichikara-instruction-004-001-0000027-001",
  "text": "「雷がゴゴロ鳴り出した」のゴゴロは擬音語で「白黒日は家であること」のこのころは擬音語である。\\n\\nこのように音が同じで直つて擬音語でも擬音語でもある事例を「ゴゴロ【でかき】」以外でつづけてください。",
  "output": "「擬音語（擬声語）」とは生体（動物）の発する音の音のすることを指し、次の鳴き声（ワウワン）や振動音（ドカーン）などがそれに該当します。\\n\\n一方、擬音語は言葉の音や文字・振動を表す音の音を指し、雷の音（かみかみ）や雷の音（かみかみ）などがそれに該当します。\\n\\n以上を前提に擬音語であり擬音語でもある事例は\\n\\n「ばらばら」\\n\\nです。擬音語の場合は雷の音（ばらばら）、擬音語の場合は雷の音（ばらばら）となります。",
  "meta": {
    "task": [
      "指示"
    ],
    "perspective": [
      "客観"
    ],
    "time-dependency": false,
    "domain": [
      "国語"
    ],
    "source-to-answer": [
      "言語知識"
    ],
    "output-type": [
      "単語"
    ],
    "text-producer": 38,
    "output-producer": 38,
    "output-reference": [
      "https://www2.ninjal.ac.jp/onomatope/column/nihongo_1.html"
    ],
    "alert-type": []
  },
}
```

さらなる発展

- 株式会社いちから（理研ベンチャー）
 - Ichikara-instructionに含まれないタイプのインストラクションの構築
 - Ichikara-instruction2: 指示追従性、COT、雑な質問など
 - Ichikara-multiturn: 対話のインストラクション
 - 2024年8月から2025年2月に構築、ライセンス提供中
 - <https://ichikara.ai>
- 情報学研究所 大規模言語モデル研究開発センター
 - AnswerCarefully（安全性インストラクション）を構築
 - 1800件の危険な事を言わないためのデータ
 - llm-jp-172Bなどに利用され大きな効果
 - 商用利用も含め公開中
 - <https://llmc.nii.ac.jp/answercarefully-dataset/>

知識グラフによる LLMの精度／信頼性の向上

LLMのハルシネーションを解決

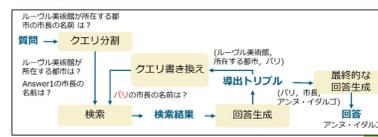
- LLMよりも信頼性の高いモジュールを組み込む
 - 外部知識の利用（RAG）
 - 様々な知識タイプが存在
- 構造化知識ベース（森羅）、文書ベースのRAGを利用
 - 森羅：AIPで構築された知識グラフ、Wikidataより高精度
 - 文書ベース：Wikipediaを利用
- マルチホップQAタスクで評価
 - 回答を出すために複数種類の知識が必要
 - 導出過程も提供することで信頼性が向上
 - 例「2024年MLBでワールドチャンピオンになったチームの主砲は誰？」
= 「2024年MLBのワールドチャンピオン」 + 「そのチームの主砲」

精度とカバレッジの問題

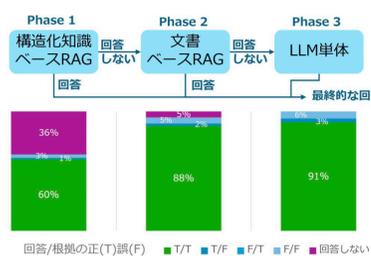
- 知識グラフ： 高精度、カバレッジは低い
- 文章RAG： 中精度、カバレッジ中程度
- LLM： 低精度、カバレッジは高い

提案手法

精度の高いものから順に適用し、答えられないものはカバレッジが高い（精度は低い）ものを適用し、回答を出していく



実験結果 (JEMHopQAデータ)



実験結果（詳細）

	LLM単体	RAG Wikidata	RAG 森羅	文書 RAG	段階的 RAG
TT	69%	43%	60%	83%	91%
TF	11%	4%	1%	3%	3%
FT	1%	0%	0%	0%	0%
FF	19%	7%	3%	6%	6%
N/A	0%	47%	36%	8%	0%

表1: 検索対象データ別 RAGの精度

TT: 回答も導出過程も正解 TF: 回答は正解だが導出過程が間違え
FT: 回答は間違えだが、導出過程は正解 FF: 回答も導出過程も間違え

TF（偽正解）：例：2人の有名人の年齢の比較
誕生日の情報は間違っているが、比較結果は合っている など