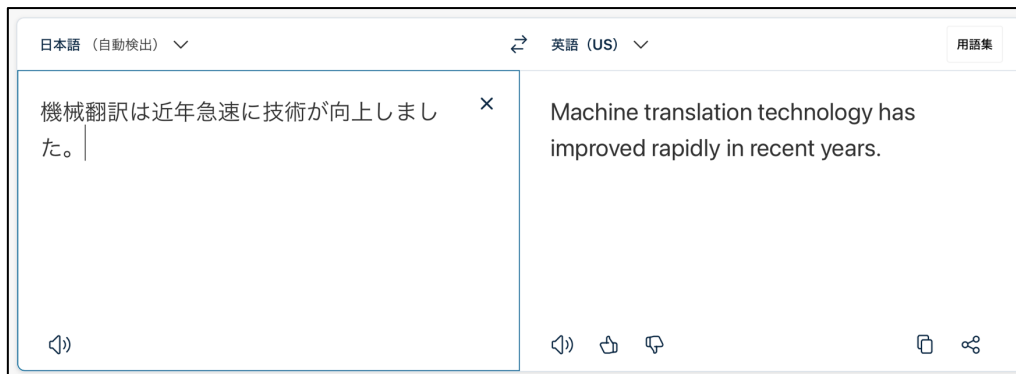


自然言語処理とは

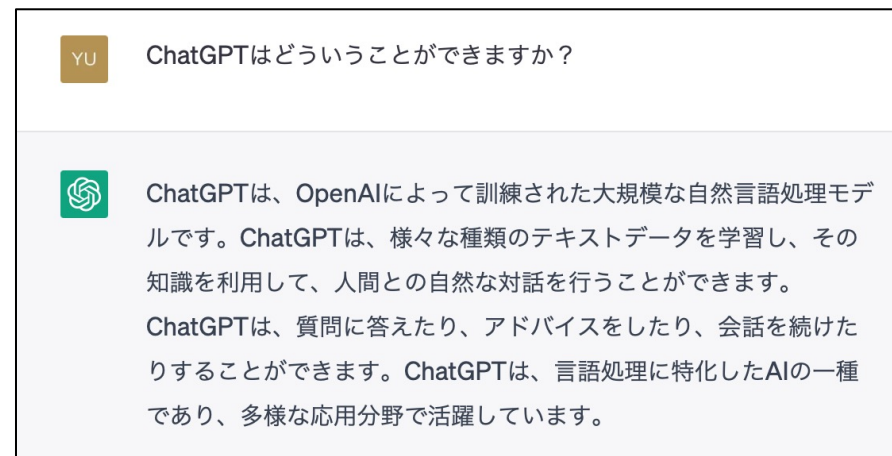
- 人間の言語能力（自然言語を理解・生成する能力）の計算モデルを研究
 - 自然言語の意味理解、推論、思考、コミュニケーション
 - 自然言語の数理モデル、解析技術、応用技術
- 自然言語処理技術を応用して人間の言語活動を助ける
- 入力または出力が自然言語である問題すべてが対象

機械翻訳



DeepL

対話システム



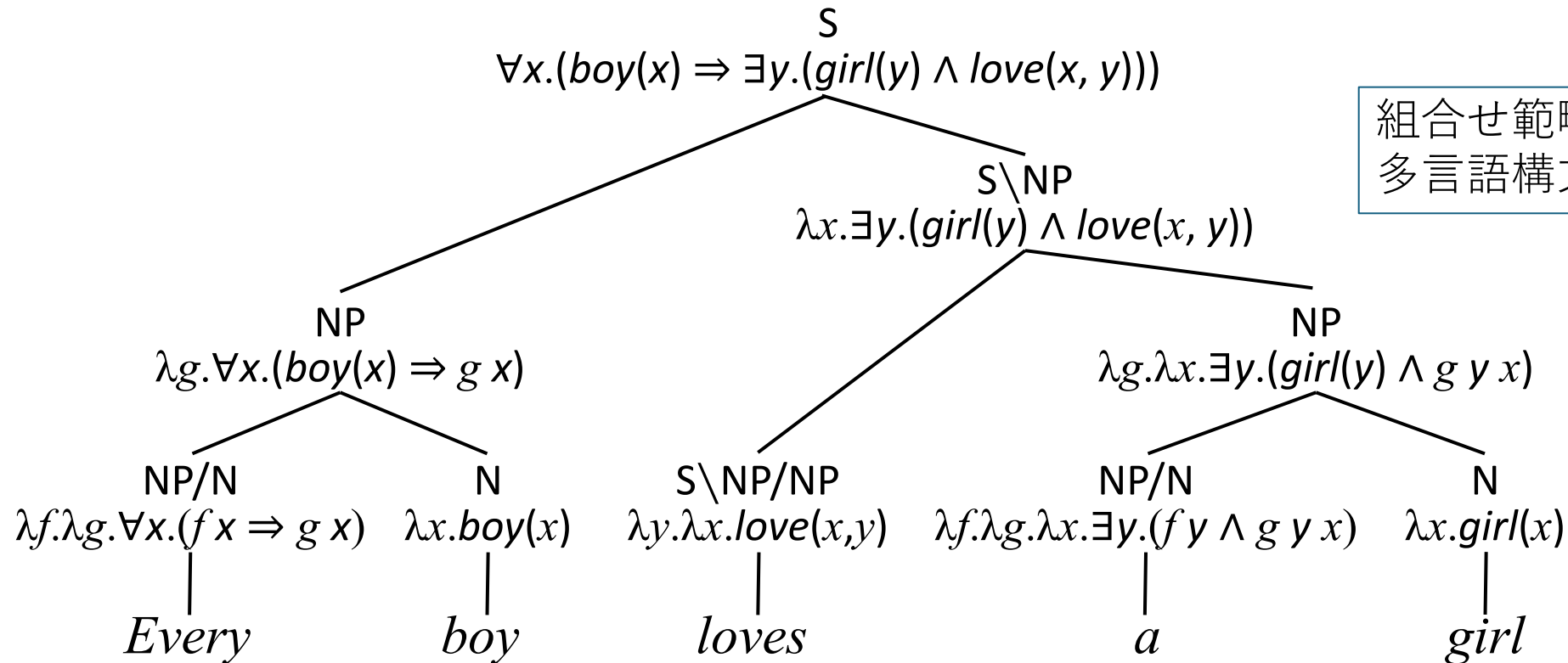
ChatGPT

研究室の研究内容

- 自然言語の構造や意味に関する研究を中心に、自然言語の本質を追及する研究に力を入れる
- 特に以下の研究テーマにフォーカス
 - **構文・意味解析**：自然言語テキストの構造解析、意味計算
 - **大規模言語モデル**：大規模言語モデルの分析、評価、高度化、実応用
 - **グラウンディング**：自然言語の意味を実世界とつなぐ理論・技術
- 学生は上記に限らず自然言語に関する研究テーマを独自に設定して研究を実施
 - 具体例は「学生の研究」参照

構文・意味解析

- 自然言語テキストの構文・意味構造を計算する技術の研究
 - 自然言語の内部構造を計算する基本問題
- 構文・意味構造に基づく高度な言語モデルの研究



組合せ範疇文法による
多言語構文解析

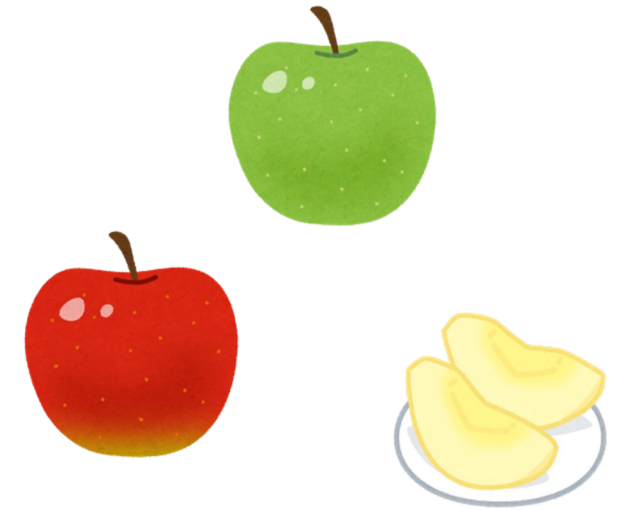
大規模言語モデル

- 人間と同レベルの言語理解能力を見せるが、そのしくみはブラックボックス
 - 大規模言語モデルの分析・評価技術の研究
- 実応用システムの開発
 - 大規模言語モデルやデータベースを組み合わせた対話システム

| | run_name | AVG ↓ | EL | FA | MC | MR | NLI | QA |
|----|---|--------|--------|--------|--------|--------|--------|--------|
| 37 | rinna/nekomata-14b | 0.6393 | 0.4533 | 0.2865 | 0.9106 | 0.6940 | 0.6240 | 0.5935 |
| 60 | augmxnt/shisa-gamma-7b-v1 | 0.6250 | 0.4210 | 0.2261 | 0.9196 | 0.6300 | 0.7611 | 0.5052 |
| 34 | rinna/nekomata-14b-instruction | 0.6070 | 0.3450 | 0.2439 | 0.9267 | 0.7060 | 0.5342 | 0.5813 |
| 65 | lightblue/karasu-7B | 0.5736 | 0.4390 | 0.2550 | 0.8534 | 0.6040 | 0.4579 | 0.4988 |
| 27 | lightblue/karasu-7B-chat-plus-unleashed | 0.5694 | 0.4035 | 0.2586 | 0.8722 | 0.6160 | 0.4382 | 0.4918 |
| 35 | lightblue/karasu-7B-chat-plus | 0.5681 | 0.4075 | 0.2541 | 0.8677 | 0.6000 | 0.4517 | 0.4928 |
| 49 | lightblue/karasu-7B-chat | 0.5670 | 0.4047 | 0.2543 | 0.8847 | 0.5800 | 0.4501 | 0.4945 |
| 79 | llm-jp/llm-jp-13b-instruct-full-jaster-dolly-oasst-v1.0 | 0.5555 | 0.3376 | 0.1578 | 0.8490 | 0.1680 | 0.9088 | 0.5433 |
| 58 | stabilityai/japanese-stablelm-instruct-gamma-7b | 0.5529 | 0.4158 | 0.2437 | 0.8597 | 0.5840 | 0.3601 | 0.5149 |
| 52 | stabilityai/japanese-stablelm-base-gamma-7b | 0.5521 | 0.4107 | 0.2534 | 0.8311 | 0.6340 | 0.3547 | 0.5012 |
| 63 | NTQAI/chatntq-ja-7b-v1.0 | 0.5482 | 0.4173 | 0.2652 | 0.7775 | 0.6120 | 0.3712 | 0.5186 |

グラウンディング

- 自然言語（頭の中）と実世界（外界）をつなぐ
 - 画像、数値データ、データベース、地理情報、etc.
 - 「リンゴ」って何？
 - 「霞ヶ関」に行きたい
- グラウンディングのための意味表現や解析技術を研究

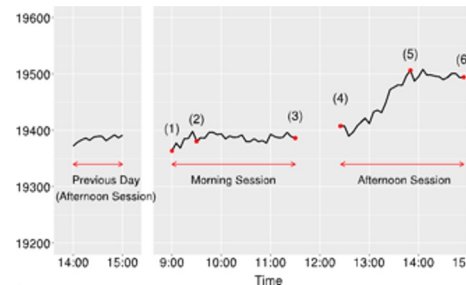


A cat is playing with a ball.



画像

Nikkei continues to fall. The closing price of the morning session decreases by 5 yen.



数値データ

Where was the author of Edo Hyakkei born?



知識グラフ

学生の研究

- 自然言語に関する幅広い研究を行なっているが、自然言語の本質にせまる基礎的研究を推奨
- 自然言語解析の基礎理論・技術
 - 多言語構文解析
 - 構文構造に基づく言語モデル
 - 議論構造の形式理論
 - 自動文章生成の制御手法
- 自然言語の情報科学的分析
 - 構文構造の言語普遍的性質の分析
 - 創発言語の分析
- 大規模言語モデル
 - 大規模言語モデルの推論能力の分析
 - 目的指向対話システムの開発
- グラウンディング
 - 画像に対する質問応答
 - 自然言語の動作への意味付け
 - 自然言語テキストにおける数式の解析