What is Natural Language Processing?

- Explore computational models of human linguistic ability (understanding/generating natural language)
 - Semantic understanding, reasoning, thinking, communication
 - Formal theory and technologies for natural language understanding
- Assist human activity by NLP technologies
- Target any problem where input and/or output is natural language

Machine translation

日本語(自動検出) 🗸	, ≁	英語 (US) 🗸	用語集	
機械翻訳は近年急速に技術が向上しまし た。	×	Machine translation technology has improved rapidly in recent years.		
م»		(1) 4 P	∂ ≪	

Dialogue system

YU	ChatGPTはどういうことができますか?
S	ChatGPTは、OpenAlによって訓練された大規模な自然言語処理モデ ルです。ChatGPTは、様々な種類のテキストデータを学習し、その 知識を利用して、人間との自然な対話を行うことができます。 ChatGPTは、質問に答えたり、アドバイスをしたり、会話を続けた りすることができます。ChatGPTは、言語処理に特化したAlの一種 であり、多様な応用分野で活躍しています。

DeepL

Research in our group

- Fundamental theories and technologies for natural language processing with particular interests in:
 - Syntactic/semantic parsing: Computing structure of sentences and their semantics
 - Large language models: analysis, evaluation, improvement, and application of LLMs
 - Grounding: ground natural language semantics on the real world
- Students research their own topics on natural language, not limited to the above
 - Refer to "Students' research" for examples

Syntactic/semantic parsing

- Research on parsing sentences into syntactic/semantic structures
 - Fundamental problem to compute the internal structure of language
- Research on syntax/semantics-aware language models



Large language models (LLMs)

- LLMs show human-level language understanding but the internal mechanism is a black box
- \rightarrow Technologies for analysis and evaluation of LLMs
- Development of LLM-based applications
 - Dialogue systems to integrate LLMs and databases

	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

Ilm-jp-eval: evaluation framework for Japanese LLMs

Grounding

Connect natural language (our brain) to the real world (outside)



- Images, numerical data, databases, geographic info, etc.
- What is "apple"?
- I'd like to visit "Hongo"!
- Develop semantic representations and technologies for grounding

A cat is playing with a ball.



Images

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



Where was the author of Edo Hyakkei born?



Students' research

- Students' research is on broad topics about natural language, while it is recommended to explore the nature of natural language rather than focusing on applications
- <u>Theories/technologies for NLP</u>
 - Multilingual syntactic parsing
 - Syntax-aware language models
 - Formal theory of argumentation structure
 - Controlling text generation
- <u>Analyzing natural language using</u> <u>computer science</u>
 - Analyzing universal properties of natural language syntax
 - Analyzing emergent languages

- <u>Large language models</u>
 - Analysis of knowledge and inference ability
 - Evaluation and Analysis of language understanding abilities
 - Goal-oriented dialogue systems
- <u>Grounding</u>
 - Grounding text into movements
 - Analyzing math formulas in natural language text
 - Speech language model