



COMP 3361

Natural Language Processing

Spring 2024

Logistics

- Location: KB 132
- Meetings: Tuesday 9:30 am - 10:20 am and Friday 9:30 am - 11:20 am
- Instructor: Tao Yu (<https://taoyds.github.io/>)
- Office hours: Wednesday 4 - 5 pm @IDS

Logistics

Course website: <https://taoyds.github.io/courses/comp3361>

- We will maintain the website for schedule, lecture slides, reading lists, grading policies, etc
- Only submit your reports on Moodle.

Logistics

Slack: https://join.slack.com/t/slack-fdv4728/shared_invite/zt-2asgddr0h-6wIXbRndwKhBw2lX2~ZrJQ

- We will use Slack as the primary mode of communication. DM me on Slack instead of emails.
 - Answer any questions about lectures, assignments, grading, and so on
 - Share random thoughts, highlight interesting papers, brag about cool finding there.
- Join Slack via the invitation link above.

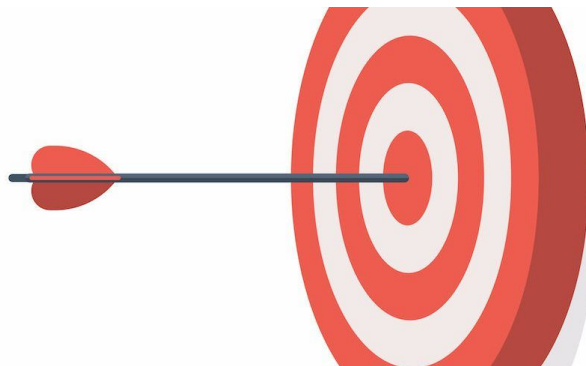


Course prerequisites

- COMP3314 or COMP3340; and MATH1853
- Familiarity with deep learning and machine learning
- Familiarity with Python programming
- Helpful: exposure to AI assistants such as ChatGPT

Course goals

- Understand core techniques and modern advances in NLP, especially in the era of large language models.
- Design, implement, and test NLP systems based on large language models.



Components and grading

- **Assignments: 40%**
 - ~2 assignments, 20% for each
- **Course project: 30%**
 - More guidelines will be announced soon
- **In-class exam: 25%**
- **Class participation: 5%**

Policy on ChatGPT, Copilot, and other AI assistants

- This course emphasizes understanding the capabilities and limitations of these AI systems, and there's no better way to do that than by using them!

Collaboration with these systems is allowed, treating them as collaborators in the problem-solving process. However, **Using them to substantially complete assignments will be considered a violation of the Honor Code.**

Class readings

- Readings from textbook chapters, blogs, tutorials, and papers will be posted on the course website.
- You may find it useful to do these readings before lecture as preparation or after lecture to review, but you are not expected to know everything discussed in the textbook if it isn't covered in lecture.
- Paper readings are intended to supplement the course material if you are interested in diving deeper on particular topics.

Topics and Schedule (Tentative)

- Introduction and NLP model basics
- Large language models (LLMs)
- NLP applications
- Advanced LLM topics

Introduction and NLP model basics

- Word embeddings
- Text Classification and Language Modeling
- Sequence-to-Sequence, Attention, Transformers

Large language models (LLMs)

- LLM pretraining
- LLM Prompting, in-context learning
- LLM evaluation, data, and benchmarking
- Instruction tuning for LLMs
- LLM alignment/RLHF

NLP applications

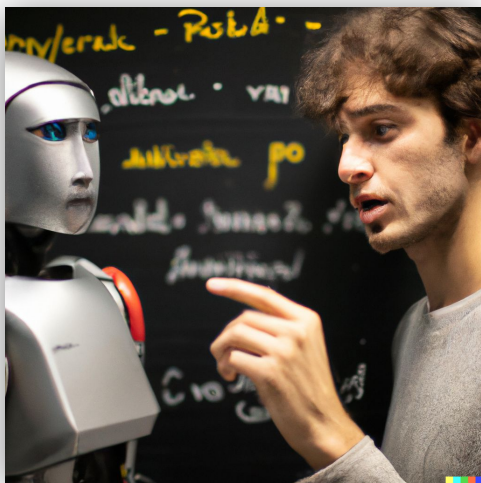
- Question answering, reasoning
- Text generation
- Semantic parsing, code generation
- LM agent, language grounding

Advanced LLM topics

- Robustness, interpretability, explainability of LLMs
- Bias, toxicity, and privacy in LLMs
- Parameter-efficient LM tuning
- Efficient LLM methods and Infrastructure
- Multimodal LM, language in robotics, and embodied interaction

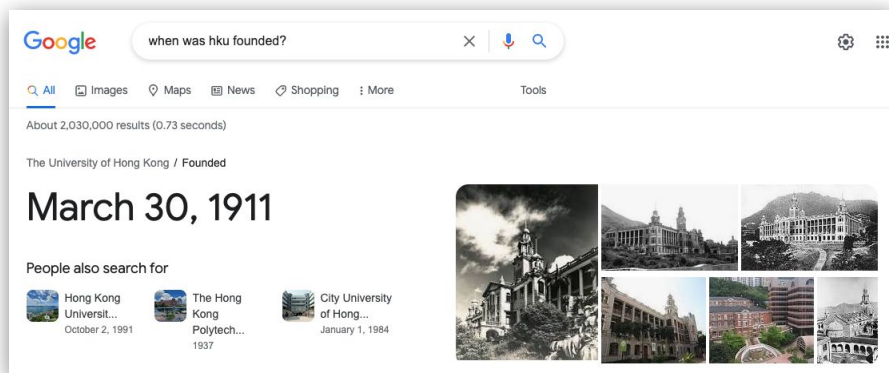
What is NLP? Wait, what is language?

- **Language** is the abstraction of the real world!
- **Natural Language Processing (NLP)** aims to teach computers human languages a computational perspective.



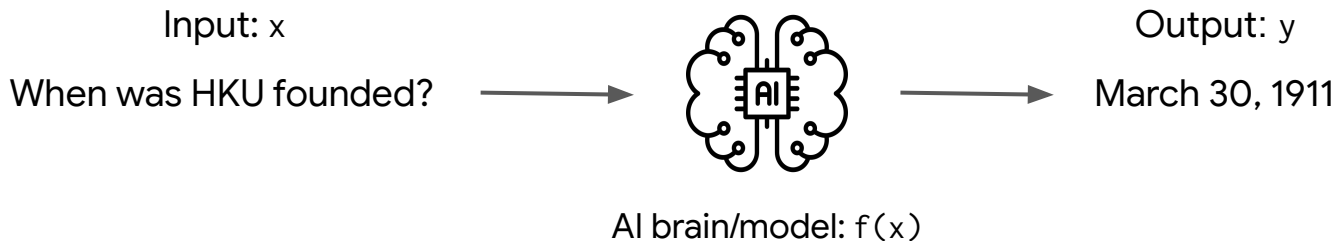
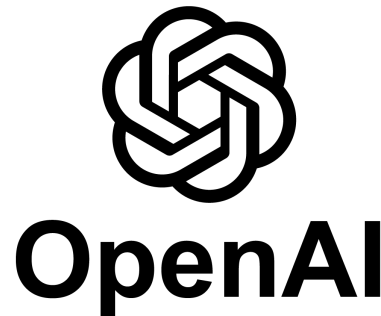
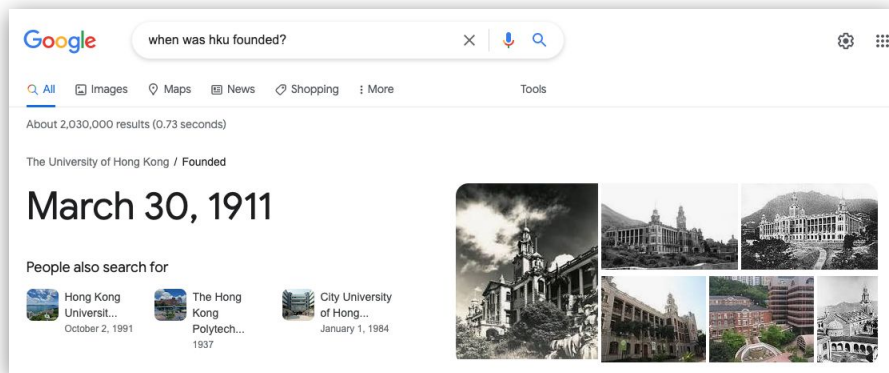
About NLP: teaching **computers** **human** languages

- NLP in real world applications
 - Q&A / IR - Google search



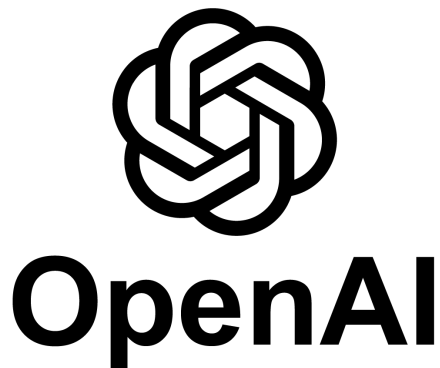
About NLP: teaching **computers** **human** languages

- NLP in real world applications
 - Q&A / IR - Google search



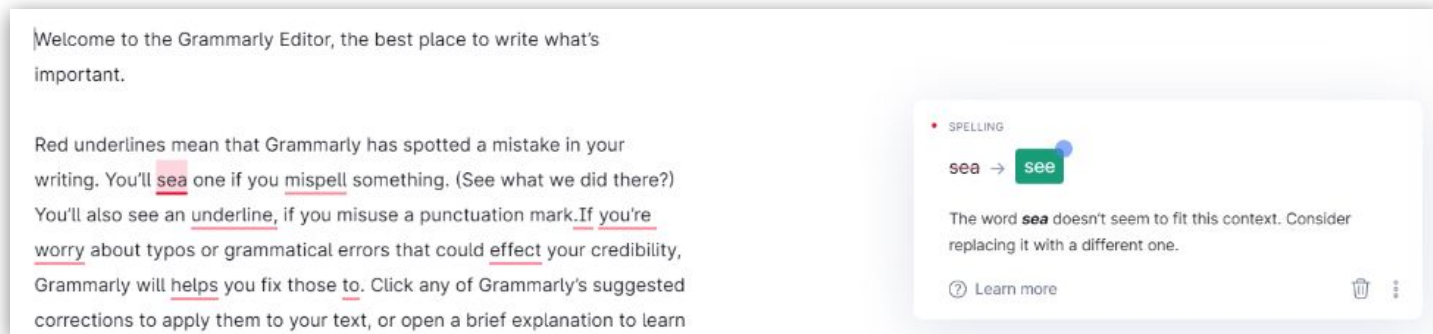
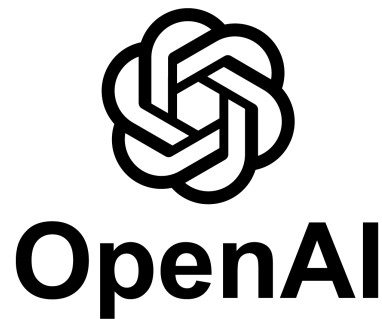
About NLP: teaching **computers** **human** languages

- NLP in real world applications
 - Q&A / IR - Google search
 - Dialogs - Apple Siri / Amazon Alexa

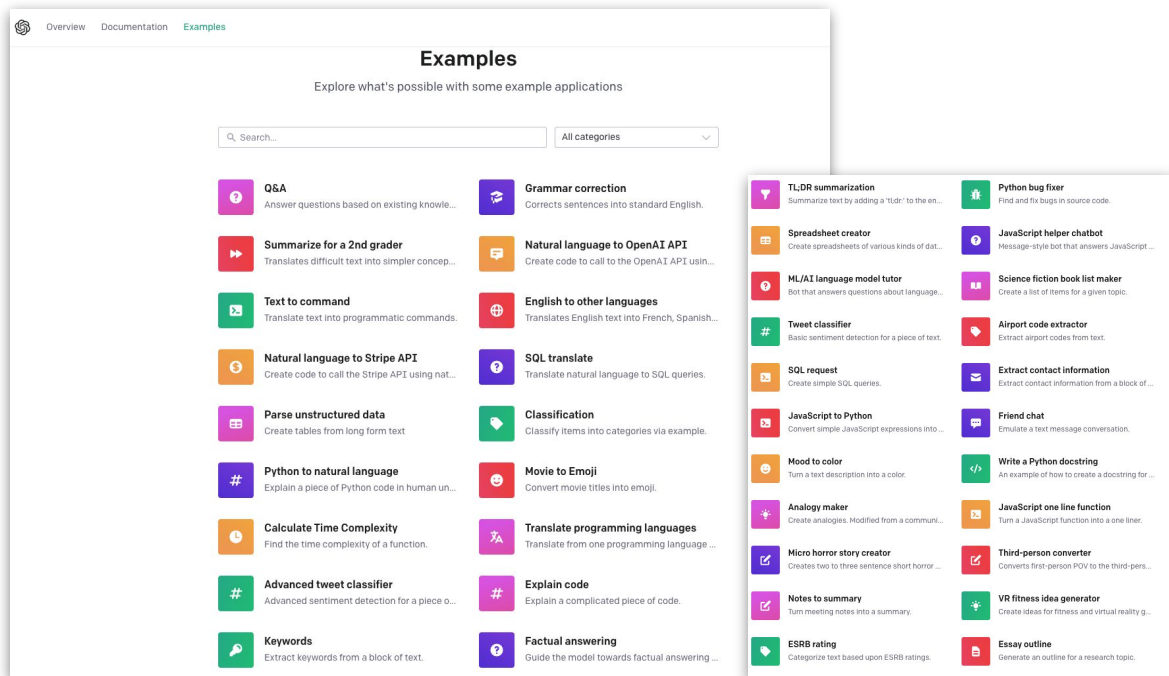


About NLP: teaching **computers** **human** languages


- NLP in real world applications
 - Q&A / IR - Google search
 - Dialogs - Apple Siri / Amazon Alexa
 - Grammar checking (Grammarly), summarization, sentiment analysis ...



What ChatGPT can do?



Q&A example with ChatGPT

 **Q&A**

Answers Generation Conversation

Answer questions based on existing knowledge.

Prompt

I am a highly intelligent question answering bot. If you ask me a question that is rooted in truth, I will give you the answer. If you ask me a question that is nonsense, trickery, or has no clear answer, I will respond with "Unknown".

Q: What is human life expectancy in the United States?
A: Human life expectancy in the United States is 78 years.

Q: Who was president of the United States in 1955?
A: Dwight D. Eisenhower was president of the United States in 1955.

Q: Which party did he belong to?
A: He belonged to the Republican Party.

Q: What is the square root of banana?
A: Unknown

Q: How does a telescope work?
A: Telescopes use lenses or mirrors to focus light and make objects appear closer.

Q: Where were the 1992 Olympics held?
A: The 1992 Olympics were held in Barcelona, Spain.

Q: How many squigs are in a bonk?
A: Unknown

Q: Where is the Valley of Kings?
A:

Sample response

The Valley of Kings is located in Luxor, Egypt.

API request

```
python ▾ Copy
1  import os
2  import openai
3
4  openai.api_key = os.getenv("OPENAI_API_KEY")
5
6  response = openai.Completion.create(
7      engine="text-davinci-002",
8      prompt="I am a highly intelligent question answering bot",
9      temperature=0,
10     max_tokens=100,
11     top_p=1,
12     frequency_penalty=0.0,
13     presence_penalty=0.0,
14     stop=["\n"]
15 )
```

More examples with ChatGPT



TL;DR summarization

Transformation

Generation

Summarize text by adding a 'tl;dr:' to the end of a text passage. It shows that the API understands how to perform a number of tasks with no instructions.

Prompt

A neutron star is the collapsed core of a massive supergiant star, which had a total mass of between 10 and 25 solar masses, possibly more if the star was especially metal-rich.[1] Neutron stars are the smallest and densest stellar objects, excluding black holes and hypothetical white holes, quark stars, and strange stars.[2] Neutron stars have a radius on the order of 10 kilometres (6.2 mi) and a mass of about 1.4 solar masses. [3] They result from the supernova explosion of a massive star, combined with gravitational collapse, that compresses the core past white dwarf star density to that of atomic nuclei.

tl;dr

Sample response

A neutron star is the collapsed core of a massive supergiant star. These ultra-dense objects are incredibly fascinating due to their strange properties and their potential for phenomena such as extreme gravitational forces and a strong magnetic field.



Interview questions

Generation

Create interview questions.


Prompt

Create a list of 8 questions for my interview with a science fiction author:

Sample response

1. What inspired you to write science fiction?
2. What themes or topics do you explore in your stories?
3. What are some of the unique elements that make science fiction stand out from other genres?
4. What challenges do you face when writing science fiction?
5. What are some of the most important aspects of a good science fiction story?
6. What do you think are the biggest challenges facing science fiction today?
7. What do you think the future of science fiction holds for readers and writers?
8. What advice would you give to aspiring science fiction authors?

Examples with ChatGPT

 **SQL request**

Transformation Generation Translation


Create simple SQL queries.

Prompt

Create a SQL request to find all users who live in California and have over 1000 credits:

Sample response

```
SELECT * FROM users WHERE state='CA' AND credits > 1000;
```

 **Python bug fixer**

Code Generation

There's a number of ways of structuring the prompt for checking for bugs. Here we add a comment suggesting that source code is buggy, and then ask codex to generate a fixed code.

Prompt

```
##### Fix bugs in the below function

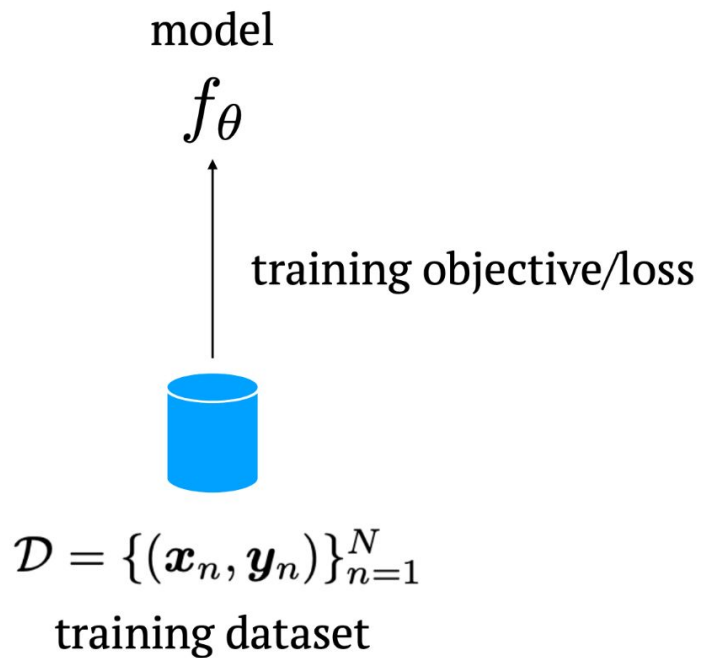
### Buggy Python
import random
a = random.randint(1,12)
b = random.randint(1,12)
for i in range(10):
    question = "What is "+a+" x "+b+"? "
    answer = input(question)
    if answer == a*b
        print (Well done!)
    else:
        print("No.")

### Fixed Python
```

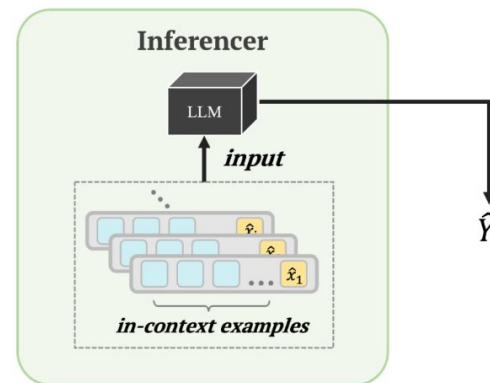
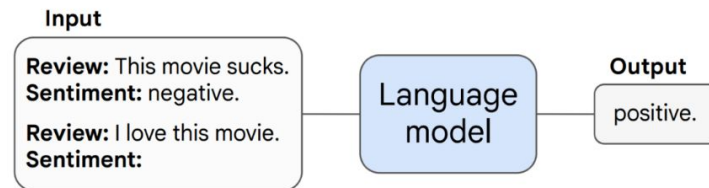
Sample response

```
import random
a = random.randint(1,12)
b = random.randint(1,12)
for i in range(10):
    question = "What is "+str(a)+" x "+str(b)+"? "
    answer = input(question)
    if answer == str(a*b):
        print ("Well done!")
    else:
```

New learning paradigm: in-context learning



Supervised Learning




In-Context Learning

Few-shot in-context learning

- **Few-shot:** In addition to the task description, the model sees **a few examples** of the task.
- **No fine-tuning**, GPT-3 doesn't update their parameters!

1	Translate English to French:	← task description
2	sea otter => loutre de mer	← examples
3	peppermint => menthe poivrée	←
4	plush girafe => girafe peluche	←
5	cheese =>	← prompt

**Q&A**

Answers Generation Conversation

Answer questions based on existing knowledge.

Prompt

I am a highly intelligent question answering bot. If you ask me a question that is rooted in truth, I will give you the answer. If you ask me a question that is nonsense, trickery, or has no clear answer, I will respond with "Unknown".

Q: What is human life expectancy in the United States?
A: Human life expectancy in the United States is 78 years.

Q: Who was president of the United States in 1955?
A: Dwight D. Eisenhower was president of the United States in 1955.

Q: Which party did he belong to?
A: He belonged to the Republican Party.

Q: What is the square root of banana?
A: Unknown

Q: How does a telescope work?
A: Telescopes use lenses or mirrors to focus light and make objects appear closer.

Q: Where were the 1992 Olympics held?
A: The 1992 Olympics were held in Barcelona, Spain.

Q: How many squigs are in a bonk?
A: Unknown

Q: Where is the Valley of Kings?
A:

Sample response

The Valley of Kings is located in Luxor, Egypt.

task description

task examples

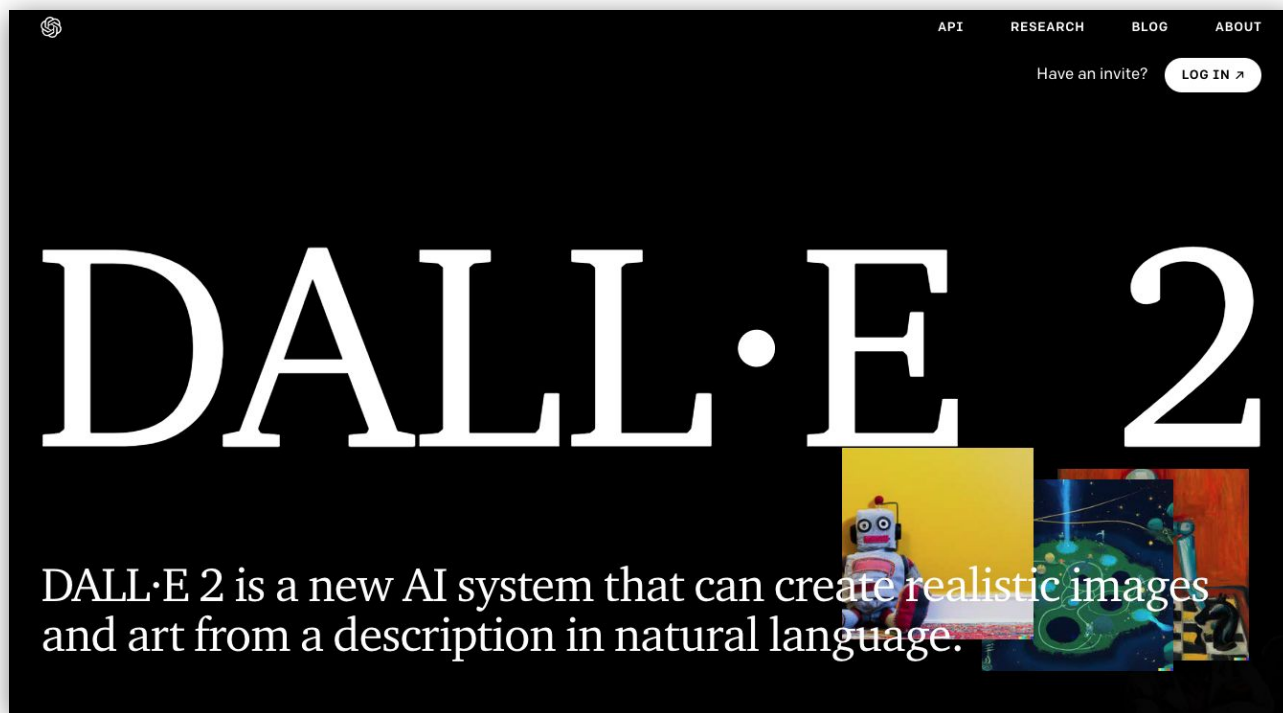
task prompt

GPT-3 outputs

About NLP: teaching **computers** **human** languages

- NLP in real world applications
 - Q&A / IR - Google search
 - Dialogs - Apple Siri / Amazon Alexa
 - Grammar checking (Grammarly), summarization, sentiment analysis ...
 - Text to images: image creation from a text description - OpenAI's DALLÉ-2

DALLE-2 demo: text to images



DALLE-2: text to images

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

mixing sparkling chemicals as mad scientists shopping for groceries working on new AI research

as kids' crayon art on the moon in the 1980s underwater with 1990s technology



DALL-E 2



Language models are powerful, but they still suffer from

- Lack of interpretability
- Inconsistency
- Limited scalability
- Restricted capabilities
- ...

