

COS IW04: Hands-on Deep Learning for Language Understanding



Spring 2022

Logistics

- **Instructor:** Danqi Chen
- **Undergraduate TAs:** TBA
- **Seminar meetings:** Mondays 11:00-12:20pm in COS402
 - Not typical lecture-style. This is independent work after all!
 - Attendance is mandatory.
 - Please follow the university COVID policy — If you can't come in-person, let me know!
- **Office hours:**
 - TA weekly office hours TBA
 - Danqi's office hour: Monday 2-3pm (appointment-based, 15 minutes each)

Logistics

Website: <https://www.cs.princeton.edu/courses/archive/spring22/cosIW04/>

COS IW04 (Spring 2022): Hands-on Deep Learning for Language Understanding

Instructor	Danqi Chen (danqi AT cs.princeton.edu)
Undergraduate TAs	TBA
Meetings	Monday 11:00am-12:20pm
Location	COS 402
Office hours	Monday 2-3pm, Zoom or COS 412 (by appointment)

Schedule

Pre-class action items should be completed by 9am on Mondays.

Week	Date	Topic	Pre-class Action Items
1	January 24	Introduction	Fill out the introduction form
2	January 31	Brainstorming project ideas	
3	February 7	Project proposal presentations	
4	February 14	Project updates	

Check the pre-class action items!



Note: I will maintain the website for schedule/useful resources, not Canvas!

Logistics

IW04: Useful Resources ☆ 📁 ☁

File Edit View Insert Format Tools Add-ons Help Last edit was 2 minutes ago

100% Normal text Arial 11 B I U A

←

Learning how to use Transfor...

Where to find NLP datasets/t...

Understanding Transformers ...

Some cool projects to explore:

1

2

3

Learning how to use Transformers/PyTorch:

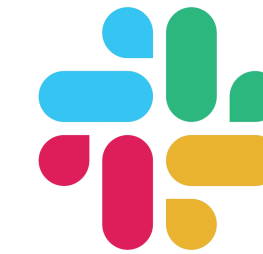
- <https://huggingface.co/docs/transformers/>
- <https://huggingface.co/course>
- I believe a good way to learn how to use Transformers is to run their examples in notebooks: <https://huggingface.co/docs/transformers/notebooks>
- If you are still not familiar with PyTorch, you need to check out some tutorials like https://pytorch.org/tutorials/beginner/nlp/pytorch_tutorial.html

Where to find NLP datasets/tasks:

- <https://huggingface.co/datasets>
- GLUE benchmark: <https://gluebenchmark.com/>

Logistics

Slack: make sure you will be added to our Slack team this week.



- We will use Slack as the primary mode of communication, so do **make sure you are notified of new messages**.
 - For important announcements (e.g., deadlines), I will write emails.
- Please feel free to use Slack to ask questions, share random musings, highlight interesting papers, brag about cool findings, even send cat photos (use **#random!**).
- I (and TAs) will be frequently on the Slack, feel free to DM us there, especially for urgent things.
- I believe that the more engaged you are on Slack, the more you can get out from the IW seminar.

Note: No Ed this semester.

What is IW about

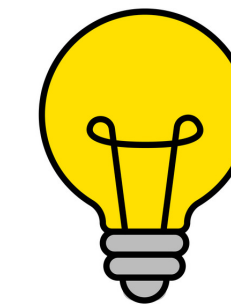
- IW = **Independent** Work
 - You have the ability to explore something that's of interest to **YOU**
 - We're here to support you in your exploration, and help scope it to be both feasible and exciting
- This IW seminar = NLP + Deep Learning (more on this later!)
- What's expected of YOU:
 - Actively participate during the seminar meetings: show up every time, engage and provide feedback to other students, **help us help you** (bring questions, concerns, progress updates)
 - Reach out if you need additional help: Slack & office hours
 - Follow all IW deadlines: <https://www.cs.princeton.edu/ugrad/independent-work/important-steps-and-deadlines>
 - Produce a great project

The semester in phases

Week	Date	Topic
1	January 24	Introduction
2	January 31	Brainstorming project ideas
3	February 7	Project proposal presentations
4	February 14	Project updates
	February 17	Deadline for a written project proposal (11:59pm)

Week 1-4: Brainstorm, literature review, finalize ideas

Week 4: deadline for written project proposal



The semester in phases

5	February 21	Project updates
6	February 28	Project updates
	March 5	Deadline for the checkpoint form (11:59pm)

Week 5-6: Initial setup (data collection, tool survey, etc)

Week 6: deadline for checkpoint form



The semester in phases

7	March 7	Spring Recess (no meeting)
8	March 14	Project updates
9	March 21	Project updates
10	March 28	Project updates
11	April 4	In-class oral presentation
12	April 11	In-class oral presentation
	April 13	Deadline for submitting video presentation (11:59pm)
13	April 18	Wrap-up meeting
	April 20	Deadline for final report (11:59pm)

Week 7-11: work hard on your project!

Week 12-13: finalize results, write report, presentations

Week 12: deadline for video presentation

Week 13: deadline for final report

Additional notes

- AB students: If you already have a proposal, you are encouraged to stick to your plan and start your project early!
- You may work in a team of 2, if the project can be split so that each student has their own semester-size piece of the project. If you are not sure, talk to me!
- All the meetings (except today) will be discussion-based and presentation-based
 - Everyone is expected to do a progress update on a weekly basis (~8 minutes each)
 - What have you done in the past week?
 - Do you have any questions?
 - Show some demos!
 - Update your Google slides before the class (before 9am)
 - I may send you some concrete instructions week by week (e.g., exploring your dataset!)

The rest of the meeting

- Introductions
- What is NLP about?
- Example of project ideas
- Useful resources

Introductions

- Who are you :)
 - Is there anything that you want the group to know about you?
- What topics in NLP are you most excited about?
- Any ML-related experiences?
- Do you have any potential ideas about this IW project already? It is okay if you haven't thought about it yet!



What is NLP about?

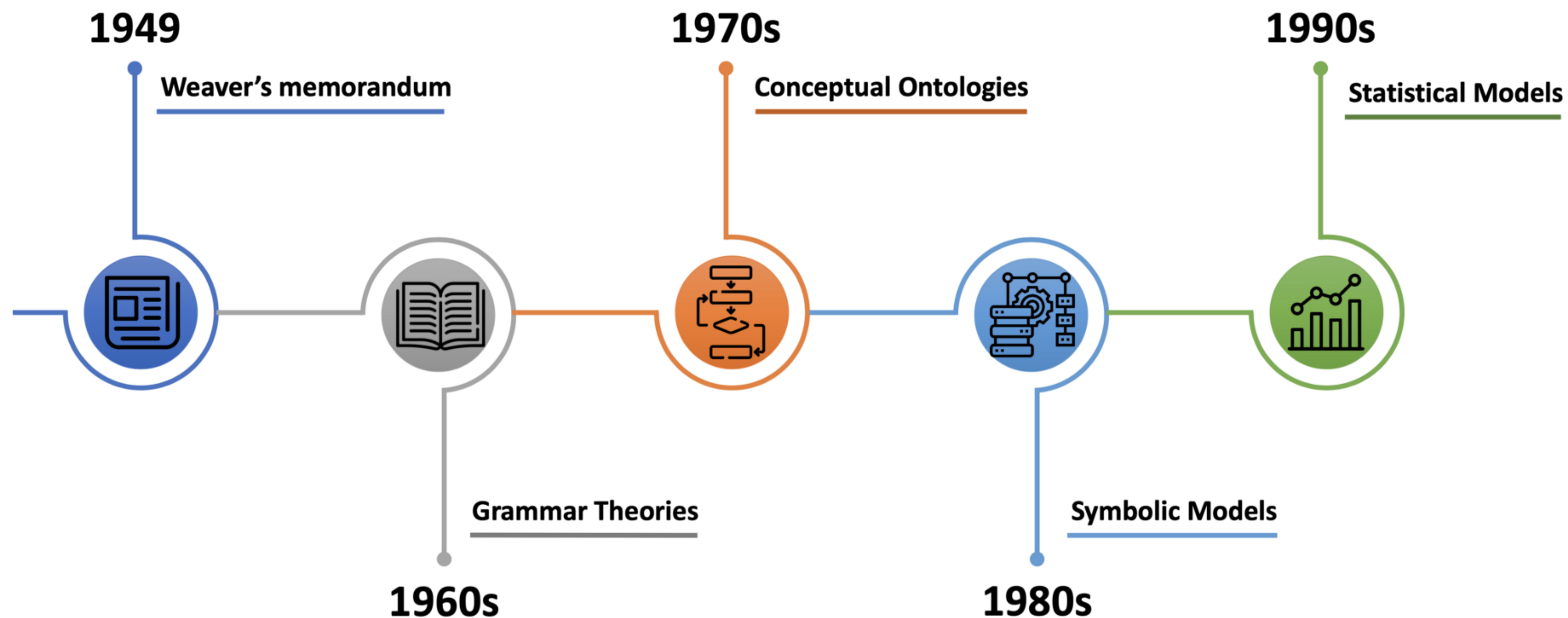
- Natural language processing (NLP) is a branch of artificial intelligence that helps computers **understand, interpret and manipulate human language**.
 - We teach computers how to read and how to write 🔍 ✎
 - What human languages do you want to work on? :)
- NLP is an extremely **rapidly-evolving** field

One of my major goals for this IW seminar! ↘

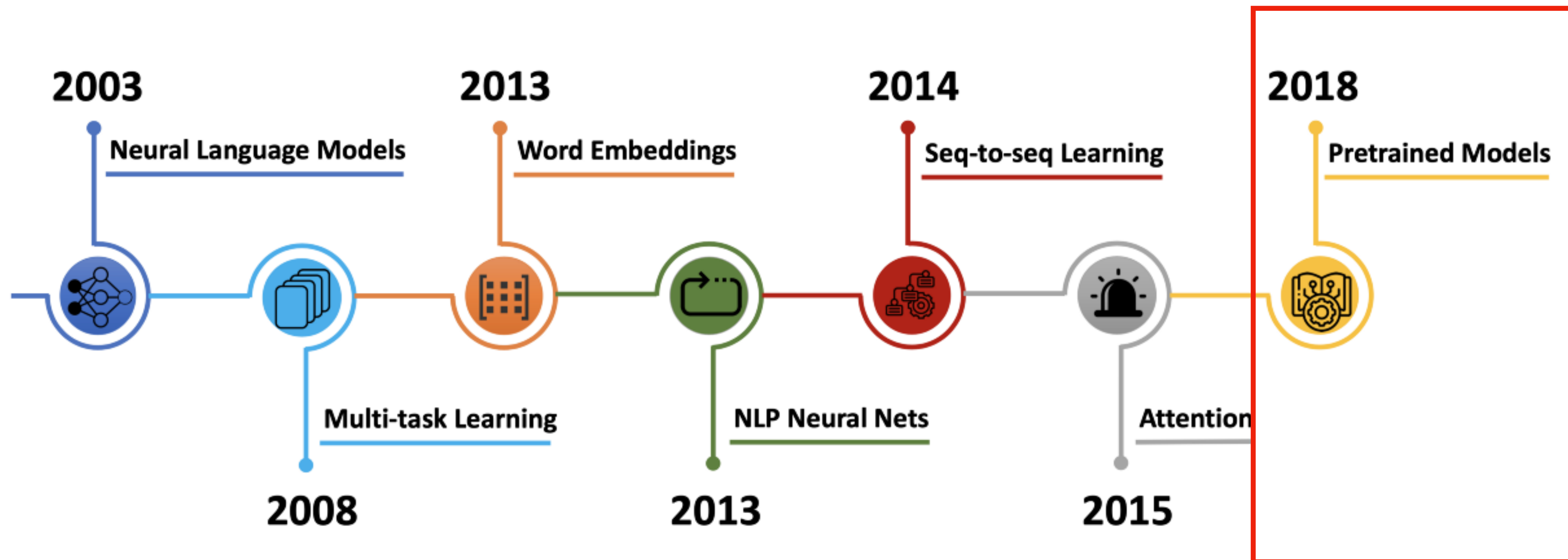
 - I highly encourage you to work with **state-of-the-art NLP models (e.g., BERT/RoBERTa/T5)**
 - They are easy to use and simply much more powerful
 - It is okay if you don't understand all the details behind them (we will catch up on this!)



A brief history of NLP

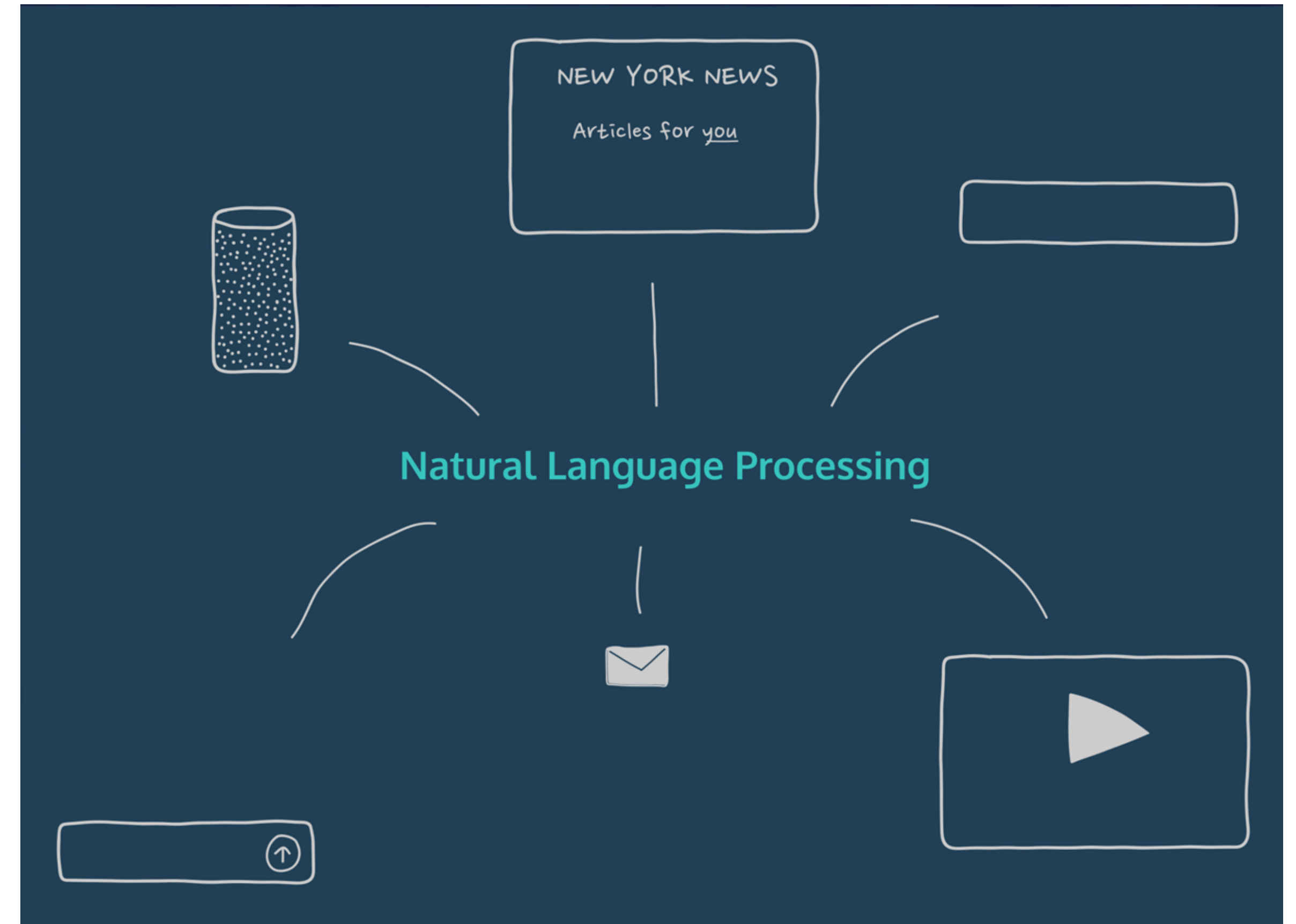
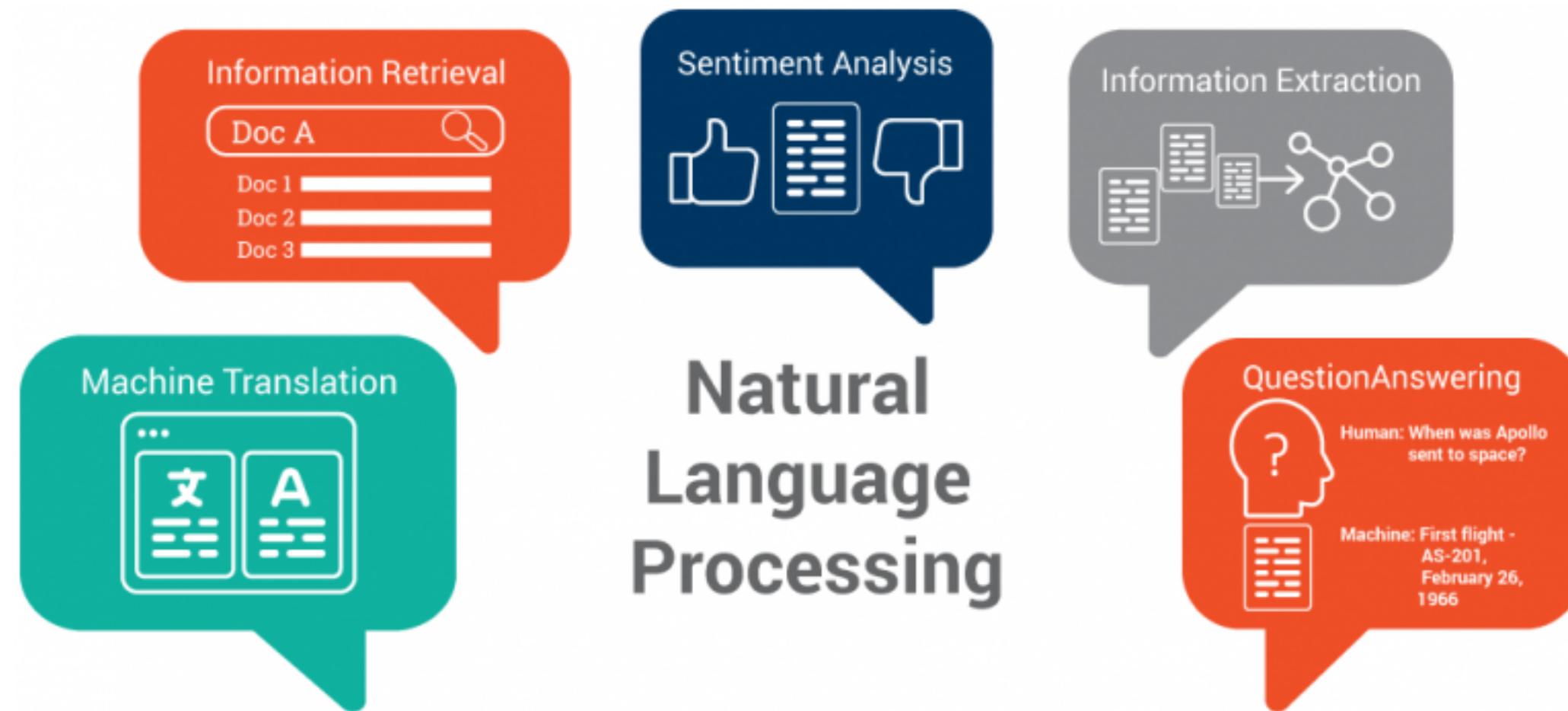


A brief history of NLP



<https://medium.com/@antoine.louis/a-brief-history-of-natural-language-processing-part-2-f5e575e8e37>

What are NLP tasks?



What are NLP tasks?

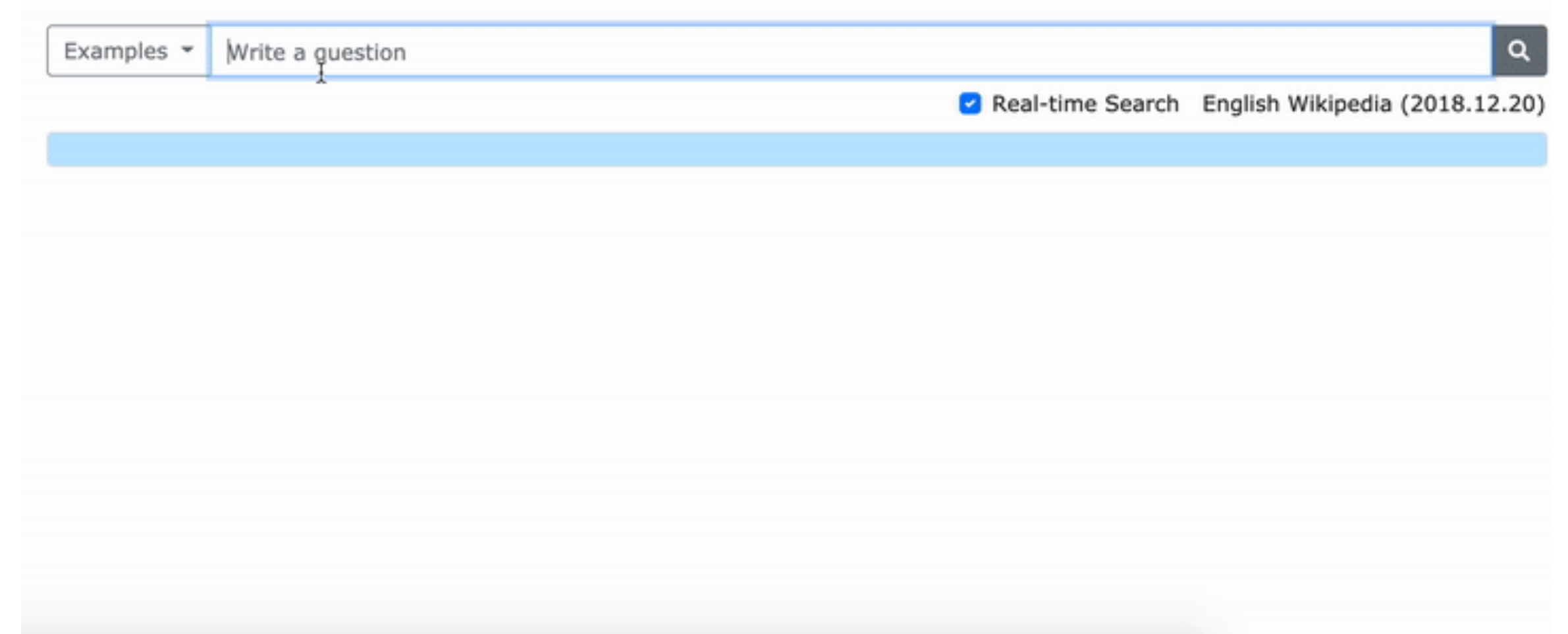
Question answering

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, *Il milione* (or, *The Million*, known in English as the *Travels of Marco Polo*), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge **through contact with Persian traders** since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: **through contact with Persian traders**

(reading comprehension)

A screenshot of a web-based question-answering interface. At the top, there is a search bar with a dropdown menu labeled "Examples" and a text input field containing "Write a question". To the right of the input field is a magnifying glass icon. Below the search bar, there is a blue bar with a checkmark icon and the text "Real-time Search English Wikipedia (2018.12.20)". The main area of the interface is a large, empty light blue rectangle, likely for displaying the answer to the question.

(open-domain QA)

What are NLP tasks?

Text summarization

Another popular pre-trained model



Source Document (abbreviated)	BART Summary
The researchers examined three types of coral in reefs off the coast of Fiji ... The researchers found when fish were plentiful, they would eat algae and seaweed off the corals, which appeared to leave them more resistant to the bacterium <i>Vibrio coralliilyticus</i> , a bacterium associated with bleaching. The researchers suggested the algae, like warming temperatures, might render the corals' chemical defenses less effective, and the fish were protecting the coral by removing the algae.	Fisheries off the coast of Fiji are protecting coral reefs from the effects of global warming, according to a study in the journal Science.
Sacoolas, who has immunity as a diplomat's wife, was involved in a traffic collision ... Prime Minister Johnson was questioned about the case while speaking to the press at a hospital in Watford. He said, "I hope that Anne Sacoolas will come back ... if we can't resolve it then of course I will be raising it myself personally with the White House."	Boris Johnson has said he will raise the issue of US diplomat Anne Sacoolas' diplomatic immunity with the White House.
According to Syrian state media, government forces began deploying into previously SDF controlled territory yesterday. ... On October 6, US President Donald Trump and Turkish President Recep Tayyip Erdoan spoke on the phone. Then both nations issued statements speaking of an imminent incursion into northeast Syria On Wednesday, Turkey began a military offensive with airstrikes followed by a ground invasion.	Syrian government forces have entered territory held by the US-backed Syrian Democratic Forces (SDF) in response to Turkey's incursion into the region.

What are NLP tasks?

Machine translation

SOURCE

作为一名艺术家，联系对我来说是非常重要的。通过我的艺术作品我试着阐明人类不是与自然分隔开而是每一件事都是互相联系的。大约10年前我第一次去了南极洲，我也第一次看到了冰山。我感到敬畏。我的心快速地砰动，头晕目眩，试着理解在我面前的这到底是什么。在我身边的冰山浮出水面几乎200英尺。我只能感到很奇怪 这就是一片雪花覆盖在另一片雪花，年复一年形成的。冰山的形成是当它们从冰川断裂开或者从冰架上断裂开。每个冰山都有它们自己的独特个性。它们与其周边的环境 和情况的互动具有一个鲜明的方式。有些冰山拒绝妥协 坚持到底，而另一些冰山就不能忍受 在一时剧烈激情喷涌下就水崩冰裂。当你看到冰山，很容易就想到它们都是孤立的，它们是独立的，单独一体的，更像是我们人类有时候对自身的看法。但现实远不止这个。随着冰山融化，我呼吸到它古老的气味。随着冰山融化，它释放了富有矿物质的淡水 它们滋养了万物。我着手拍摄这些冰山 好似我在拍摄我祖先的肖像，了解到在这些个别的时刻 冰山原是以那样方式存在 但再也不会像那样存在了。当它们融化时，这绝不是死亡；也绝不是一个终结，而是通往生生不息之路的一个延续。我拍摄过的冰山，有些冰是非常年轻--几千年年龄。有些冰 超过十万年。我想给大家展示的最后图片 是我在格陵兰岛的 Kekertsuatsiak上拍摄的一个冰山。这是一个非常难得的机会 大家实际上得以见证 一个冰山翻滚。所以这就如图所示。在左边你能看到一个小船。这是一个约15英尺的船。我想让你注意 冰山的形状 它在水面上的变形。在这儿你看到它开始翻滚，小船移动到另一边，一个男人站在那里。这是一个平均尺寸的格陵兰冰山。它浮出水面大约有120英尺高 或者40米高。这视频是实时拍摄的。就像这冰山，它们展示给大家的是其个性的不同方面。谢谢。

**mBART25
SENT-MT**

As an artist, connection is very important to me. And through my artwork, I try to illustrate that humans are not separate from nature, but that everything is connected to each other. I went to Antarctica for the first time about 10 years ago, and I saw the iceberg for the first time. I felt awe. And my heart was pounding, my head was pounding, trying to figure out what was going on in front of me. And the iceberg that was right next to me was floating almost 200 feet above the surface of the water, and I just couldn't help but feel that this was a snowflake covering another snowflake, forming over and over and over again. The formation of glaciers is when they break off from glaciers, or break off from ice shelves. Every iceberg has its own unique personality. They interact in a very distinctive way with the environment around them and with the circumstances around them. Some glaciers refused to compromise and insisted, and some glaciers couldn't stand the heat of passion as it poured down the ice. When you look at icebergs, it's easy to think that they're isolated, that they're independent, that they're separate, that they're more like the way we think about ourselves sometimes. But the reality is much more than that. And as the glacier melts, I breathe in its ancient smell. And as the glacier melts, it releases fresh water of minerals that nourish everything. I started photographing these icebergs like I was photographing my ancestors, and I learned that in these individual moments, the icebergs existed in that way, but they never existed like that again. When they melt, it's not about death; it's not about the end, it's about the continuation of a life-long path. I photographed glaciers, and some of them were very young -- thousands of years old. Some of the ice has been there for more than 100,000 years. And the last picture I want to show you is an iceberg that I photographed in Kekertsuatsiak on the island of Greenland. It's a very difficult opportunity to actually witness the rolling of an iceberg. So this is what it looks like. You can see a little boat on the left. This is a 15-foot boat. I want you to notice that the shape of the iceberg changes as it moves over the surface. And here you see it rolling, and the boat moves to the other side, and a man is standing there. This is an average size glacier in Greenland. It floats about 120 feet up or 40 meters above the surface. This video was taken in real time. And like this iceberg, they show you different aspects of their personality. Thank you.

What are NLP tasks?

Sentiment analysis

Loves the German bakeries in Sydney. Together with my imported honey it feels like home	Positive
@VivaLaLauren Mine is broken too! I miss my sidekick	Negative
Finished fixing my twitter...I had to unfollow and follow everyone again	Negative
@DinahLady I too, liked the movie! I want to buy the DVD when it comes out	Positive
@frugaldougal So sad to hear about @OscarTheCat	Negative
@Mofette brilliant! May the fourth be with you #starwarsday #starwars	Positive
Good morning thespians a bright and sunny day in UK, Spring at last	Positive
@DowneyisDOWNEY Me neither! My laptop's new, has dvd burning/ripping software but I just can't copy the files somehow!	Negative

What are NLP tasks?

Natural language inference

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

What are NLP tasks?

Fact checking

Claim: Claire Danes is wedded to an actor from England.

[wiki/Claire_Danes] [She is married to actor Hugh Dancy](#), with whom she has one child.

[wiki/Hugh_Dancy] [Hugh Michael Horace Dancy](#) (born 19 June 1975) [is an English actor](#) and model.

Verdict: SUPPORTED

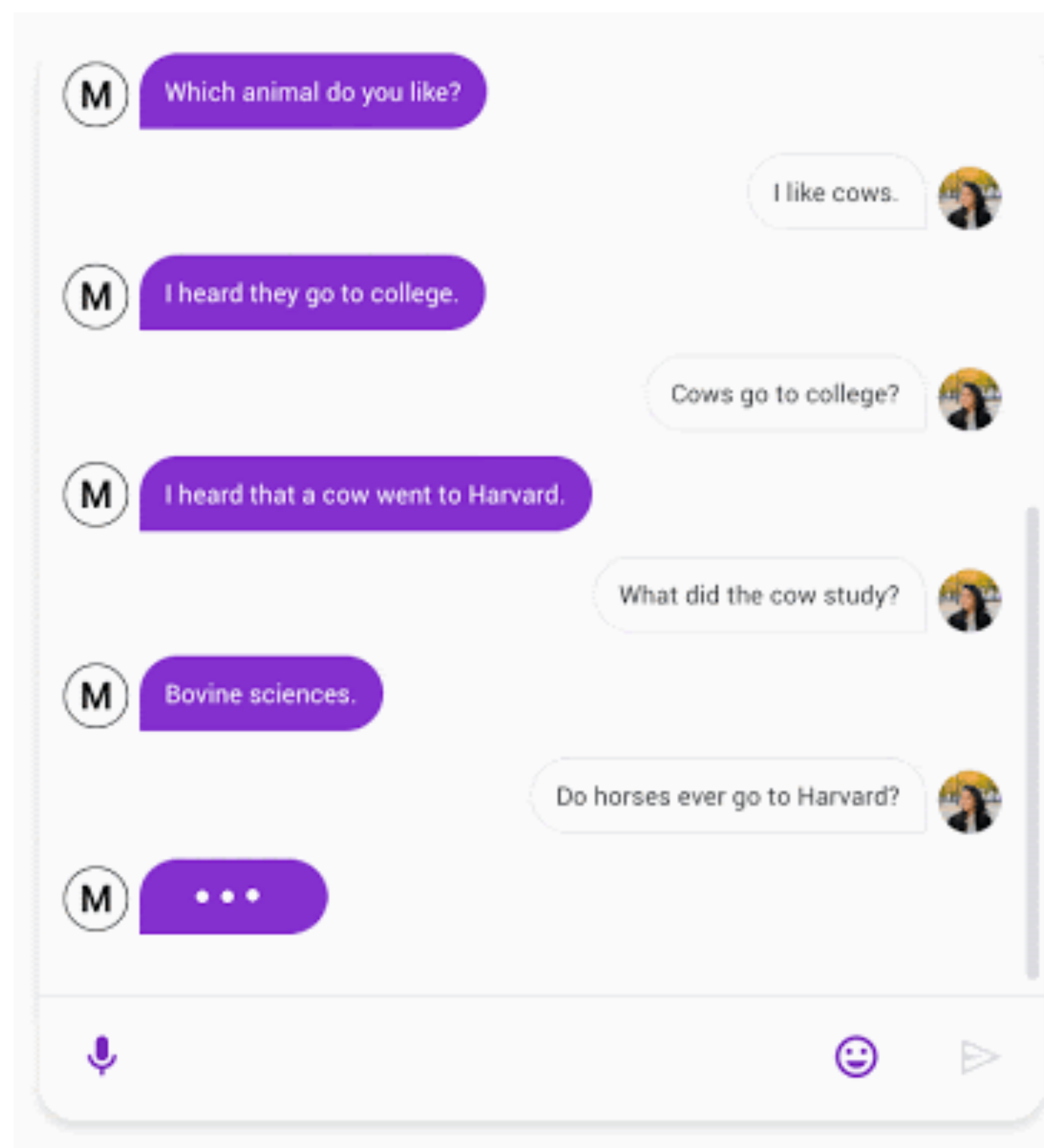
Claim: Rogue appears in Canadian comic books.

[wiki/Rogue_(comics)] [Rogue is a fictional superhero appearing in American comic books](#) published by Marvel Comics, commonly in association with the X Men .

Verdict: REFUTED

What are NLP tasks?

Dialogue



NLP benchmarks

- GLUE
- SuperGLUE
- XTREME (cross-lingual benchmarks)
- ... and many others

NLP benchmarks

Dataset	Description	Data example	Metric
CoLA	Is the sentence grammatical or ungrammatical?	"This building is than that one." = Ungrammatical	Matthews
SST-2	Is the movie review positive, negative, or neutral?	"The movie is funny , smart , visually inventive , and most of all , alive ." = .93056 (Very Positive)	Accuracy
MRPC	Is the sentence B a paraphrase of sentence A?	A) "Yesterday , Taiwan reported 35 new infections , bringing the total number of cases to 418 ." B) "The island reported another 35 probable cases yesterday , taking its total to 418 ." = A Paraphrase	Accuracy / F1
STS-B	How similar are sentences A and B?	A) "Elephants are walking down a trail." B) "A herd of elephants are walking along a trail." = 4.6 (Very Similar)	Pearson / Spearman
QQP	Are the two questions similar?	A) "How can I increase the speed of my internet connection while using a VPN?" B) "How can Internet speed be increased by hacking through DNS?" = Not Similar	Accuracy / F1
MNLI-mm	Does sentence A entail or contradict sentence B?	A) "Tourist Information offices can be very helpful." B) "Tourist Information offices are never of any help." = Contradiction	Accuracy
QNLI	Does sentence B contain the answer to the question in sentence A?	A) "What is essential for the mating of the elements that create radio waves?" B) "Antennas are required by any radio receiver or transmitter to couple its electrical connection to the electromagnetic field." = Answerable	Accuracy
RTE	Does sentence A entail sentence B?	A) "In 2003, Yunus brought the microcredit revolution to the streets of Bangladesh to support more than 50,000 beggars, whom the Grameen Bank respectfully calls Struggling Members." B) "Yunus supported more than 50,000 Struggling Members." = Entailed	Accuracy
WNLI	Sentence B replaces sentence A's ambiguous pronoun with one of the nouns - is this the correct noun?	A) "Lily spoke to Donna, breaking her concentration." B) "Lily spoke to Donna, breaking Lily's concentration." = Incorrect Referent	Accuracy

NLP benchmarks

Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI	AX
1	Microsoft D365 AI & MSR	MT-DNN-ensemble		87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9	87.9	87.4	96.0	86.3	89.0	42.8
		MT-DNN		85.1	65.4	95.6	91.2/88.1	89.6/89.0	73.7/89.9	87.9	87.4	96.0	85.7	75.3	42.8
2	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9	-
3	王玮	ALICE large ensemble (Alibaba D)		86.3	68.6	95.2	92.6/90.2	91.1/90.6	74.4/90.7	88.2	87.9	95.7	83.5	80.8	43.9
4	Stanford Hazy Research	Snorkel MeTaL		83.2	63.8	96.2	91.5/88.5	90.1/89.7	73.1/89.9	87.6	87.2	93.9	80.9	65.1	39.9
5	张倬胜	SemBERT		82.9	62.3	94.6	91.2/88.3	87.8/86.7	72.8/89.8	87.6	86.3	94.6	84.5	65.1	42.4
6	Anonymous Anonymous	BERT + BAM		82.3	61.5	95.2	91.3/88.3	88.6/87.9	72.5/89.7	86.6	85.8	93.1	80.4	65.1	40.7
7	Nitish Shirish Keskar	Span-Extractive BERT on STILTs		82.3	63.2	94.5	90.6/87.6	89.4/89.2	72.2/89.4	86.5	85.8	92.5	79.8	65.1	28.3
8	Jason Phang	BERT on STILTs		82.0	62.1	94.3	90.2/86.6	88.7/88.3	71.9/89.4	86.4	85.6	92.7	80.1	65.1	28.3
9	Jacob Devlin	BERT: 24-layers, 16-heads, 1024-l		80.5	60.5	94.9	89.3/85.4	87.6/86.5	72.1/89.3	86.7	85.9	92.7	70.1	65.1	39.6
10	Neil Houlsby	BERT + Single-task Adapters		80.2	59.2	94.3	88.7/84.3	87.3/86.1	71.5/89.4	85.4	85.0	92.4	71.6	65.1	9.2
11	Zhuohan Li	Macaron Net-base		79.7	57.6	94.0	88.4/84.4	87.5/86.3	70.8/89.0	85.4	84.5	91.6	70.5	65.1	38.7

<https://gluebenchmark.com/leaderboard>

NLP benchmarks

 **Hugging Face**

[Models](#) [Datasets](#) [Spaces](#) [Docs](#) [Solutions](#) [Pricing](#) [Log In](#) [Sign Up](#)

Task Categories

[text-classification](#) [conditional-text-generation](#) [sequence-modeling](#) [question-answering](#) [structure-prediction](#) [other](#) + 37

Tasks

[machine-translation](#) [language-modeling](#) [named-entity-recognition](#) [extractive-qa](#) [sentiment-classification](#) [summarization](#) + 215

Languages

[en](#) [es](#) [de](#) [fr](#) [pt](#) [it](#) + 187

Multilinguality

[monolingual](#) [multilingual](#) [translation](#) [other-programming-languages](#) [fa](#) [en](#) + 4

Sizes

[10K< n < 100K](#) [1K < n < 10K](#) [100K < n < 1M](#) [1M < n < 10M](#) [unknown](#) [n < 1K](#) + 15

Datasets 2,630

↑↓ Sort: Most Downloads



glue

Updated Apr 14, 2021 • ↓ 593k • ♥ 15



super_glue

Updated Nov 22, 2021 • ↓ 170k • ♥ 6



imdb

Updated Nov 5, 2021 • ↓ 111k • ♥ 3



squad

Updated Jul 19, 2021 • ↓ 76.6k • ♥ 14



tweets_hate_speech_detection

Updated Dec 6, 2021 • ↓ 45.8k • ♥ 1



anli

Updated Apr 14, 2021 • ↓ 42.4k • ♥ 3



trec

Updated Apr 26, 2021 • ↓ 37.7k • ♥ 2



common_voice

Updated Dec 6, 2021 • ↓ 269k • ♥ 20



blimp

Updated Nov 22, 2021 • ↓ 145k



wikitext

Updated Nov 9, 2021 • ↓ 94.3k • ♥ 5



wmt16

Updated Nov 30, 2020 • ↓ 48.7k • ♥ 2



squad_v2

Updated Jul 7, 2021 • ↓ 43.4k • ♥ 2



librispeech_asr

Updated Dec 6, 2021 • ↓ 42.1k • ♥ 6



xnli

Updated Nov 22, 2021 • ↓ 30.6k • ♥ 5

<https://huggingface.co/datasets>

Types of IW projects

- Scenario #1: “I have an application problem that I am really interested in. I would like to see how to apply state-of-the-art NLP techniques on it.”
 - Perfect! Make sure that you have the dataset ready in the first 2-3 weeks.
- Scenario #2: “I want to start with a state-of-the-art NLP model and I want to study its behaviors, understand its limitations, and perhaps even improve it!”
 - Perfect! There are tons of things you can do here.
 - Example: Can I make my model run faster?
 - Example: Can I break a current system? Even better, how to fix it?
 - Example: How to build a system if I only have 100 training examples?
 - Example: Is a current system biased? How to analyze it and how to fix it?
- Scenario #X: Feel free to come up with your ideas and we can always talk!

Types of IW projects

Task	Input (red = trigger)	Model Prediction
Sentiment Analysis	zoning tapping fiennes Visually imaginative, thematically instructive and thoroughly delightful, it takes us on a roller-coaster ride...	Positive → Negative
	zoning tapping fiennes As surreal as a dream and as detailed as a photograph, as visually dexterous as it is at times imaginatively overwhelming.	Positive → Negative

(Wallace et al., 2019)

Article: Super Bowl 50

Paragraph: *“Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver’s Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.”*

Question: *“What is the name of the quarterback who was 38 in Super Bowl XXXIII?”*

Original Prediction: John Elway

Prediction under adversary: Jeff Dean

(Jia and Liang, 2017)

Types of IW projects

Mention -----coref----- Mention -----coref----- Mention -----coref----- Mention
The surgeon could n't operate on his patient : it was his son !

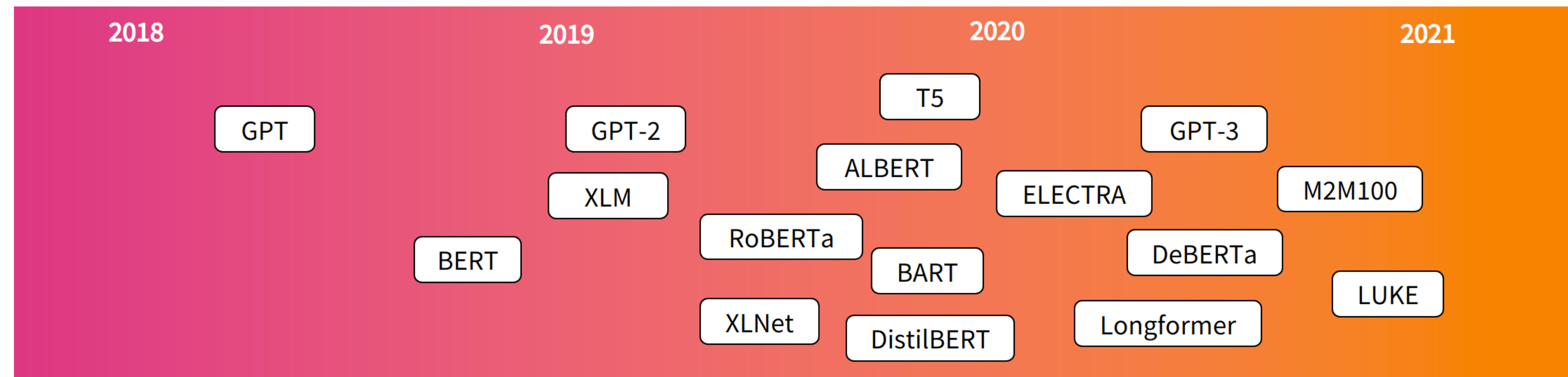
Mention -----coref----- Mention -----coref----- Mention -----coref----- Mention
The surgeon could n't operate on their patient : it was their son !

Mention -----coref----- Mention -----coref----- Mention -----coref----- Mention
The surgeon could n't operate on her patient : it was her son !

(Rudinger et al, 2018)

Transformers library 🤗

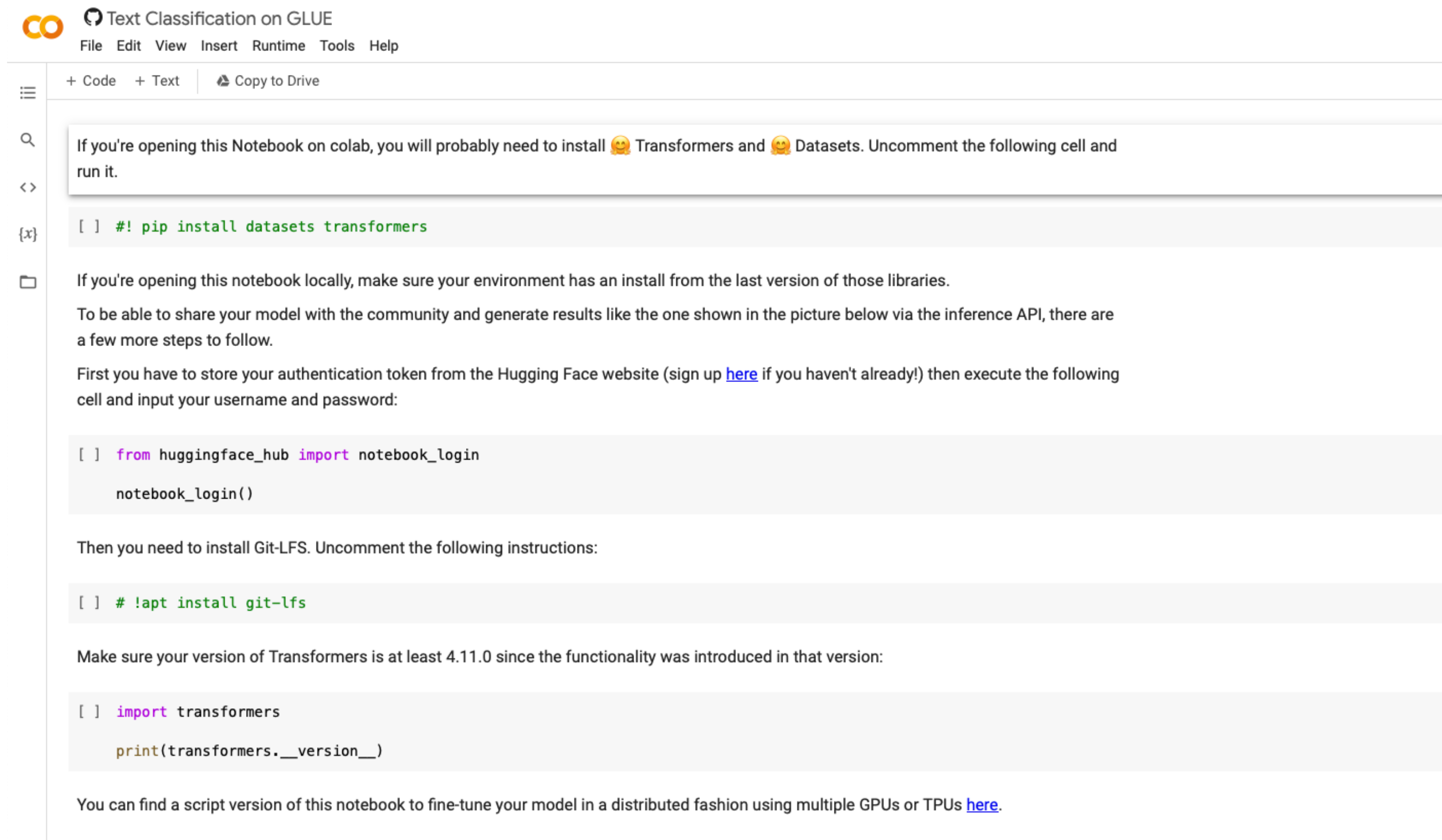
- <https://huggingface.co/docs/transformers/index>



- Transformers = a type of neural network introduced by Google researchers in 2017
 - Transformer **encoder**
 - Transformer **decoder**: usually for generation tasks
- Pre-trained models = a set of Transformer models you can download directly for your use
 - The weights are “pre-trained” instead of random initialization

Transformers library 🤗

Check out their notebooks: <https://huggingface.co/docs/transformers/notebooks>



The screenshot shows a Google Colab notebook interface. At the top, the title is "Text Classification on GLUE". Below the title is a menu bar with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". There are also buttons for "+ Code", "+ Text", and "Copy to Drive". The notebook content includes:

- A text block: "If you're opening this Notebook on colab, you will probably need to install 🤗 Transformers and 🤗 Datasets. Uncomment the following cell and run it."
- A code cell with the command:

```
[ ] #! pip install datasets transformers
```
- A text block: "If you're opening this notebook locally, make sure your environment has an install from the last version of those libraries. To be able to share your model with the community and generate results like the one shown in the picture below via the inference API, there are a few more steps to follow. First you have to store your authentication token from the Hugging Face website (sign up [here](\"#\") if you haven't already!) then execute the following cell and input your username and password:"
- A code cell with the commands:

```
[ ] from huggingface_hub import notebook_login  
  
notebook_login()
```
- A text block: "Then you need to install Git-LFS. Uncomment the following instructions:"
- A code cell with the command:

```
[ ] # !apt install git-lfs
```
- A text block: "Make sure your version of Transformers is at least 4.11.0 since the functionality was introduced in that version:"
- A code cell with the commands:

```
[ ] import transformers  
  
print(transformers.__version__)
```
- A text block: "You can find a script version of this notebook to fine-tune your model in a distributed fashion using multiple GPUs or TPUs [here](\"#\")."

https://colab.research.google.com/github/huggingface/notebooks/blob/master/examples/text_classification.ipynb

Compute resources

- Use Google Colab (you can do lots of things with 1 free GPU)
- Cloud computing: <https://researchcomputing.princeton.edu/systems/cloud-computing>
- Tiger cluster?

2021-2022 IW and Thesis Projects Schedule

Spring 2022 Single Semester Projects Schedule

Jan 13	Deadline to Select an IW Seminar or Adviser
Jan 25	Attend the "Getting Started" Information Meeting
Feb 11	 SEAS Fall Funding Deadline

You can apply for this!

Next meeting: brainstorming

- Everyone submits 2 initial ideas and we will discuss it in the meeting
 - More instructions coming soon
- I encourage everyone to start looking at the Transformers library
 - <https://huggingface.co/docs/transformers/>
 - <https://huggingface.co/course>