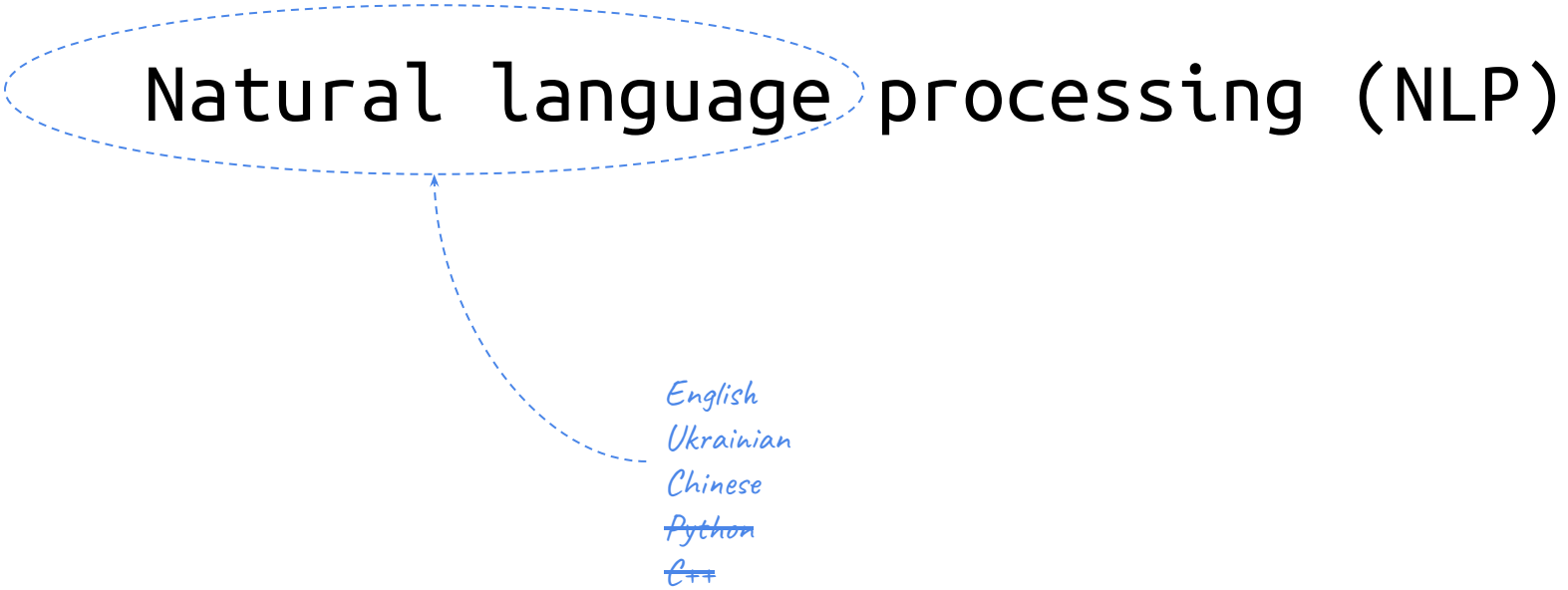
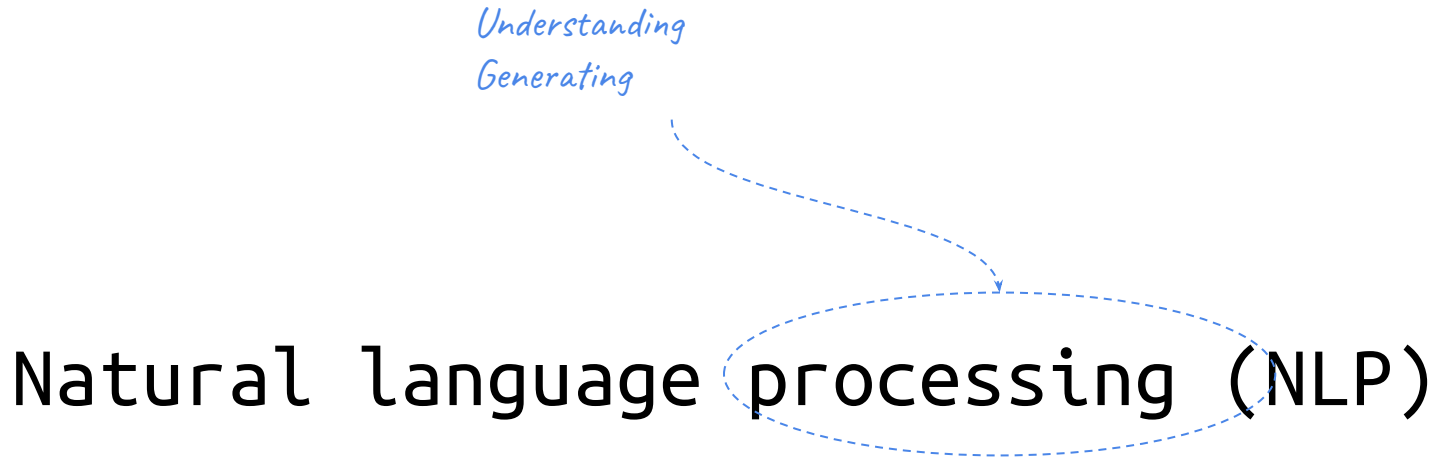
Class introduction

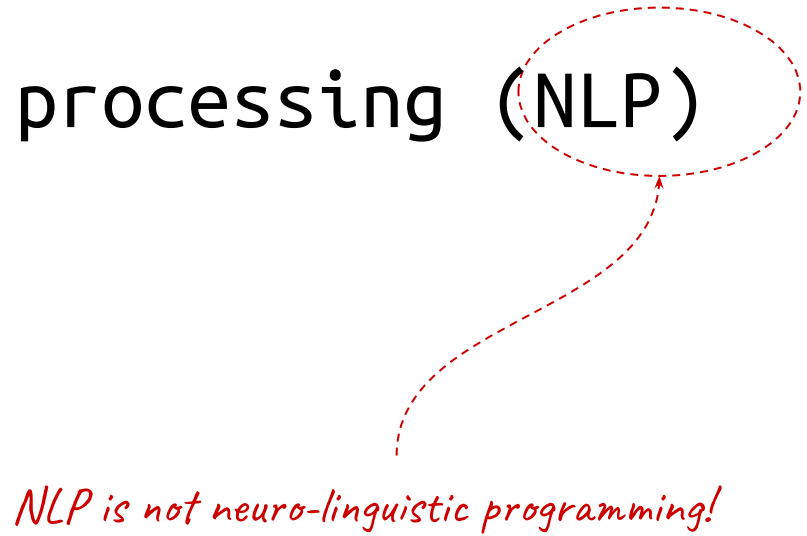
LP NLP
2025

Natural language processing (NLP)


Natural language processing (NLP)
English
Ukrainian
Chinese
~~Python~~
~~C++~~


Understanding
Generating
Natural language processing (NLP)

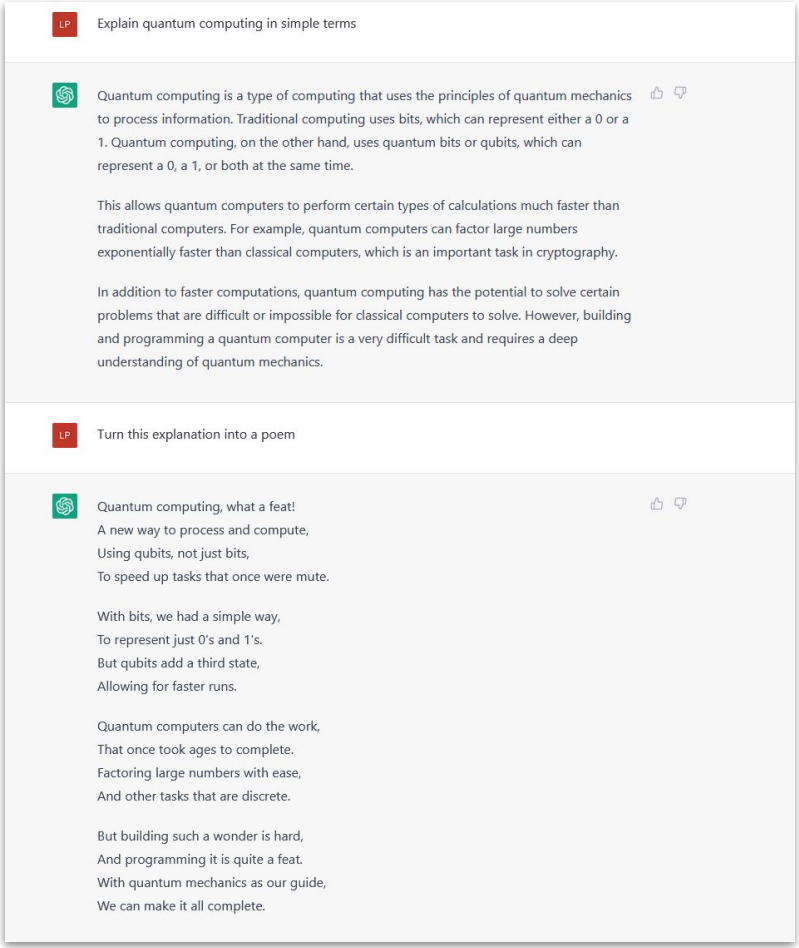
Natural language processing (NLP)



	Natural Language Processing	**Computational Linguistics**
Approach	More practical	More theoretical
Goal	Practical applications	Understand the language itself
Methods	Machine and deep learning	Rule-based and formal models
Background	Computer science and data science	Linguistics

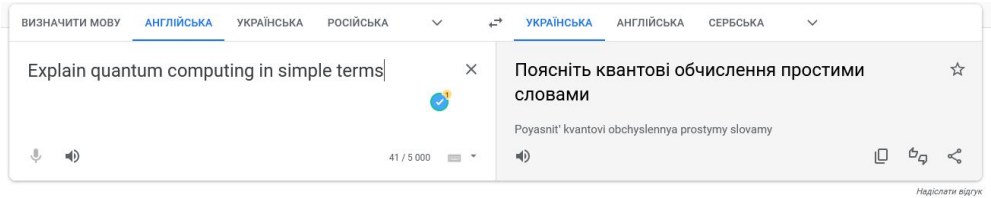
NLP Applications

Large language models / chatbots



Large language models / chatbots

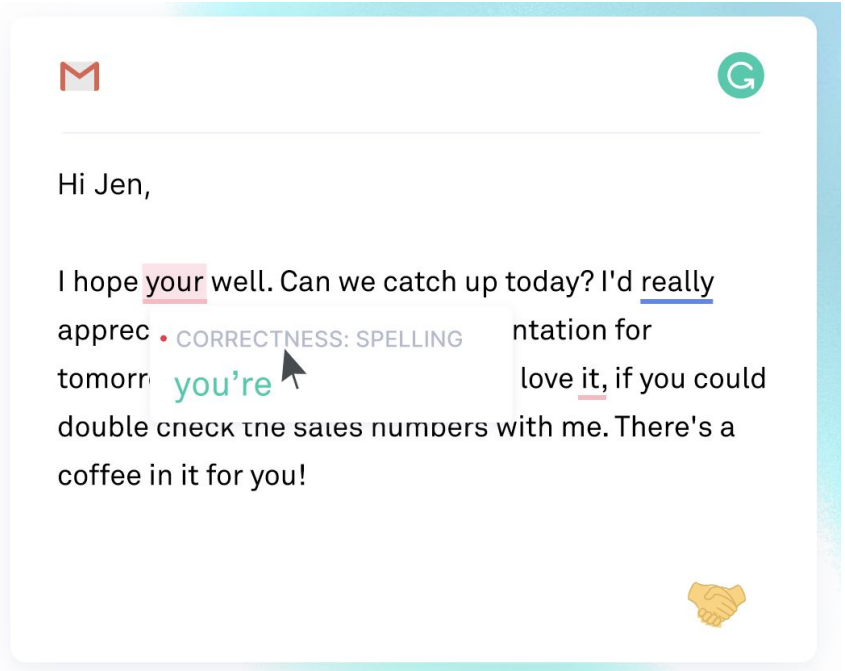
Machine translation



Large language models / chatbots

Machine translation

Grammar and spelling correction



Large language models / chatbots

Machine translation

Grammar and spelling correction

Spam detection

Question answering

Text summarization

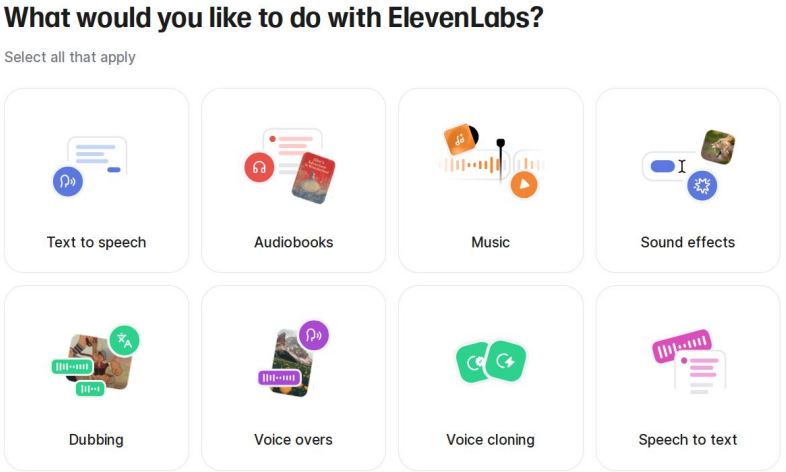
...

Speech recognition

Text-to-speech

Speech-to-speech

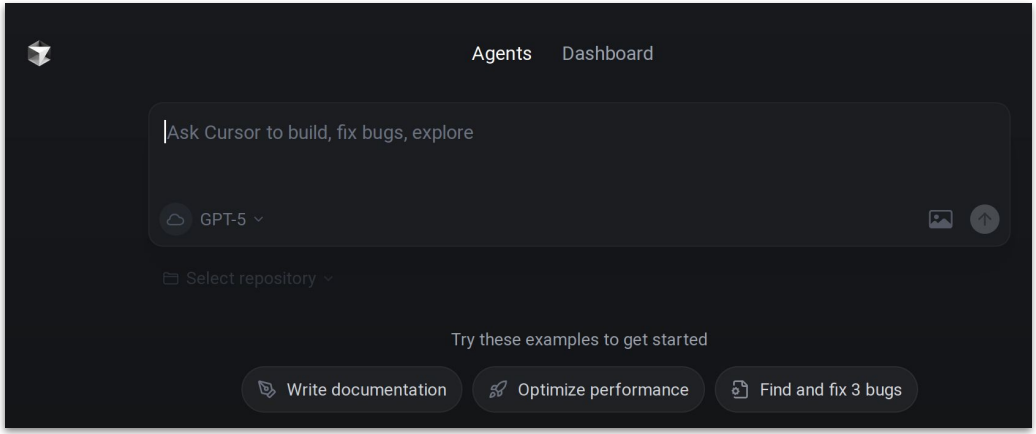
- Voice cloning
- Dubbing



Code completion

SWE-Agents

- Code generation
- Bug fixing
- Code optimization
- Documentation
- ...



Language is hard

Ambiguity

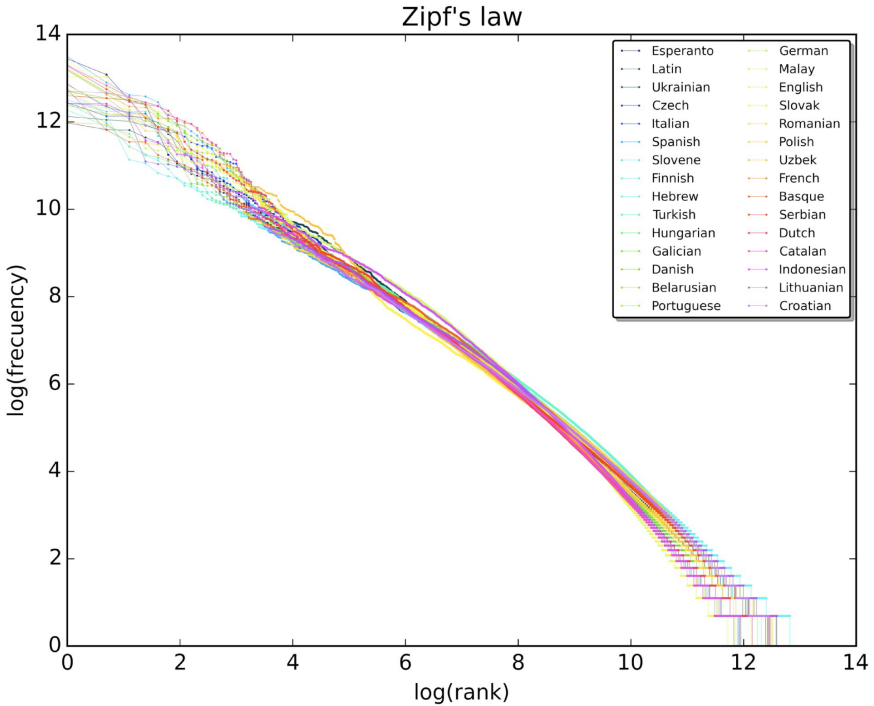
Я їм суп з куркою.

Я їм суп з другом.

Я їм суп із задоволенням.

Ambiguity

Sparsity



Ambiguity

Sparsity

Neologisms

Doomscrolling

Metaverse

NFT (Non-Fungible Token)

Рашизм

Шахед

Бавовна

Ambiguity

Sparsity

Neologisms

Over 7000 languages

~50% have no written form

New dialects emerge

Ambiguity

Sparsity

Neologisms

Over 7000 languages

Cultural references

Sarcasm, irony

...

Practical: Many applications

Research: Many open questions, path to AGI (maybe)

Many ideas are applicable to speech processing, code generation, computer vision, robotics, etc.

Exciting progress in recent years

NLP is fun!

› NLP and AI

Turing test

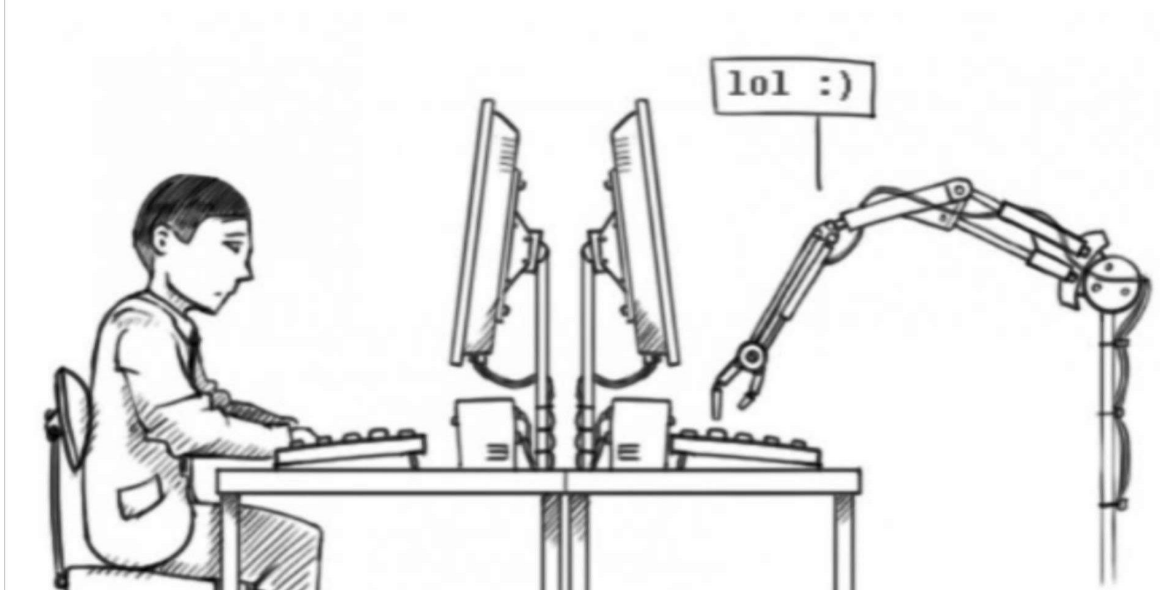


image source

Turing test

Winograd schemas

The trophy doesn't fit into the brown suitcase because it is too **small**

The trophy doesn't fit into the brown suitcase because it is too **large**

Turing test

Winograd schemas

Acquiring new knowledge

from existing texts

from dialogs

Turing test

Winograd schemas

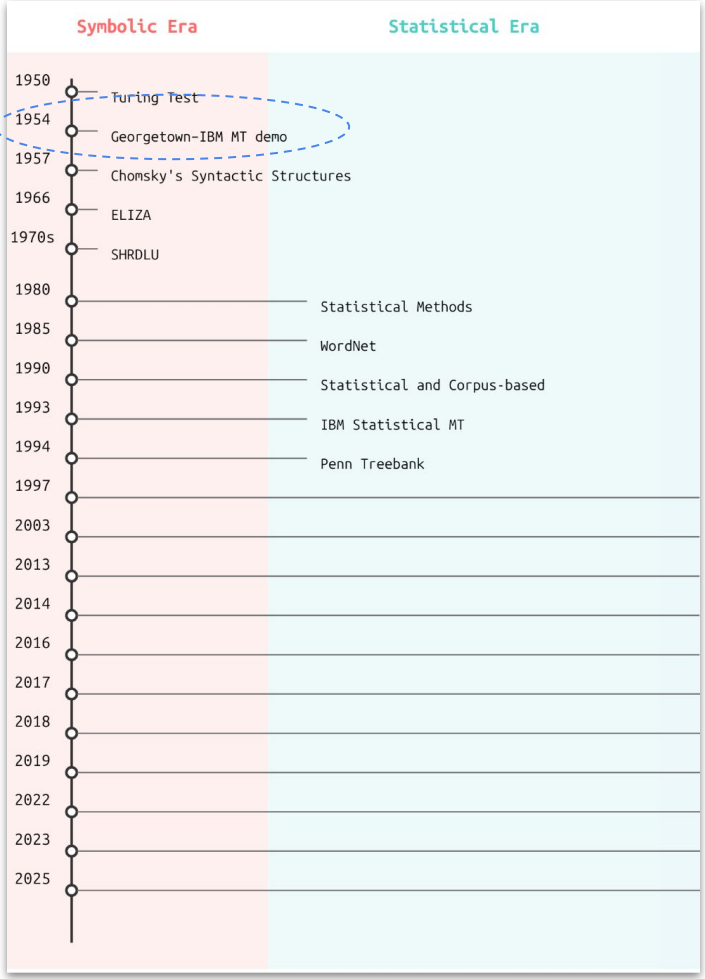
Acquiring new knowledge

Some disagree 🤷

“It would be a great achievement if AI could attain the capabilities of an orangutan, which do not include language”

- Yann LeCun

› History of NLP



Georgetown-IBM experiment

Implemented 6 rules (estimated total is ~100).

250 lexical units (words and endings)

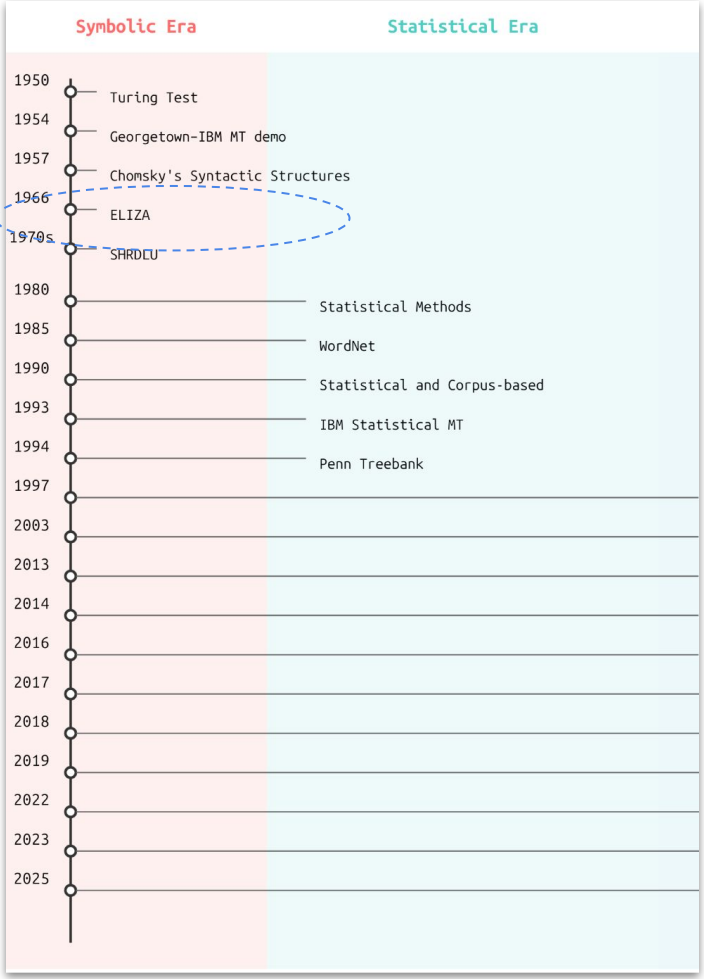
Punch cards for input

> the formulation of logic required to convert word meanings properly even in a small segment of two languages necessitates two and a half times as many instructions to the computer as are required to simulate the flight of a guided missile

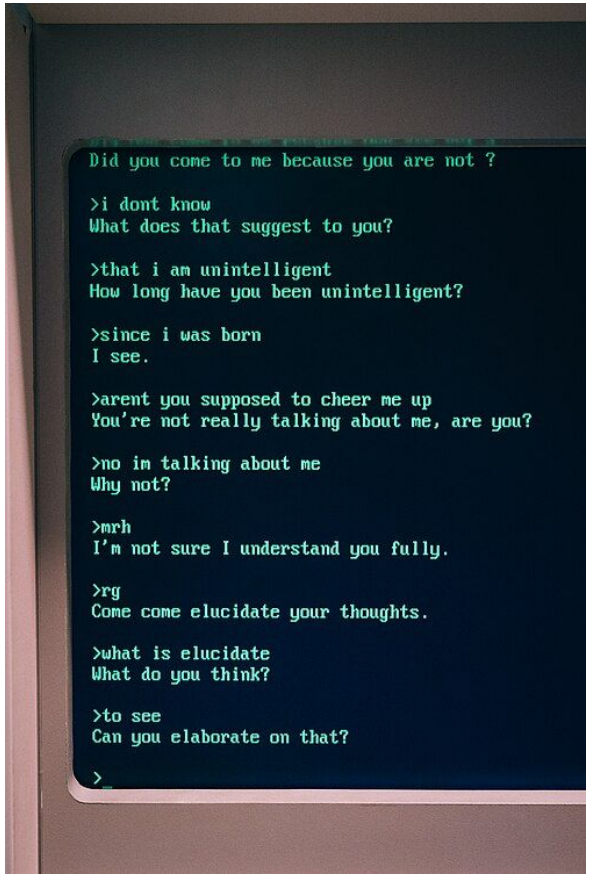
\- IBM press release

> within three or five years, machine translation could well be a solved problem

Closed in 1966


Symbolic Era
Statistical Era
1950
Turing Test
1954
Georgetown-IBM MT demo
1957
Chomsky's Syntactic Structures
1966
ELIZA
1970s
SHRDLU
1980
Statistical Methods
1985
WordNet
1990
Statistical and Corpus-based
1993
IBM Statistical MT
1994
Penn Treebank
1997
2003
2013
2014
2016
2017
2018
2019
2022
2023
2025

ELIZA


Did you come to me because you are not ?
>i dont know
What does that suggest to you?
>that i am unintelligent
How long have you been unintelligent?
>since i was born
I see.
>arent you supposed to cheer me up
You're not really talking about me, are you?
>no im talking about me
Why not?
>mrh
I'm not sure I understand you fully.
>rg
Come come elucidate your thoughts.
>what is elucidate
What do you think?
>to see
Can you elaborate on that?
>_

Logistics

Oleksiy Syvokon

Сивоконь Олексій Олегович

AI Engineer @ Zed Industries

oleksii.o.syvokon@lpnu.ua

2025 ZED INDUSTRIES

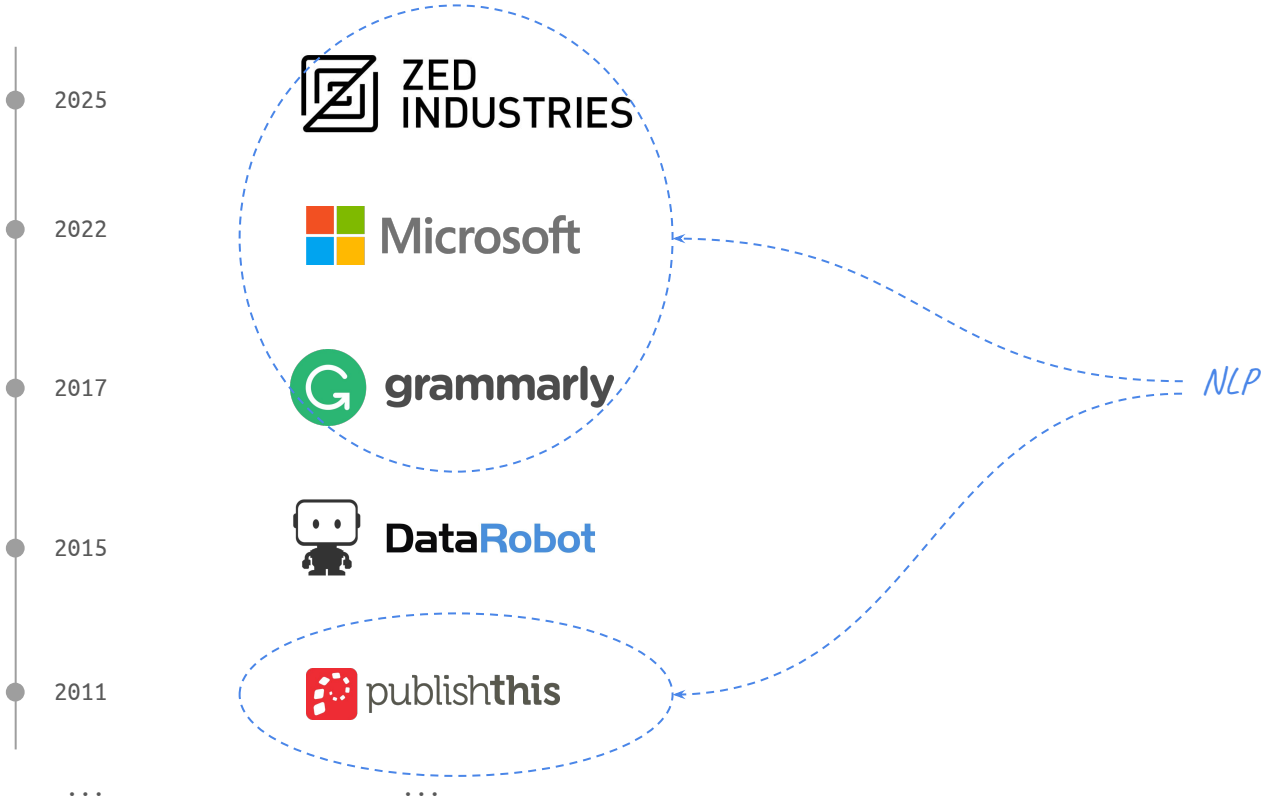
2022 Microsoft

2017 grammarly

2015 DataRobot

2011 publishthis

... ...


2025
2022
2017
2015
2011
...
...
ZED INDUSTRIES
Microsoft
grammarly
DataRobot
publishthis
NLP

Basic machine learning

Classification, regression, clustering

Overfit, regularization

Hyperparameters search

Basic machine learning

Deep learning

Feed-forward neural networks

Backpropagation

Basic machine learning

Deep learning

Software engineering

Python

Jupyter notebooks

PyTorch

Basic machine learning

Deep learning

Software engineering

Math

Basic linear algebra

Basic probability

You can:

- Solve real-world tasks
- Read papers
- Showcase your portfolio
- Pass NLP interviews

~15	lectures	
~8	labs	90
	Exam	
	Course notes	5

~15	lectures	
~8	labs	90
	Exam	
	Course notes	5

1. Text classification
2. Word embeddings
3. N-gram language models

"classic" NLP

4. Feed-forward and RNN language models
5. Transformer
6. Text generation
7. Sequence tagging
8. Sequence-to-sequence models. Machine translation
9. Transfer learning. Parameter-efficient fine-tuning
10. Large language models (LLM)

deep learning for NLP

11. Prompt engineering
12. Retrieval-augmented generation
13. Agents
14. Preference optimization
15. Modern Transformer and alternative architectures
16. Multimodality

LLMs

Rule-based NLP

Shallow ML NLP

Linguistics and computational linguistics

What’s next

Questions?

Join our Discord server

Text classification

https://discord.gg/xXHnAry6