

Class introduction

Natural language processing (NLP)

Natural language processing (NLP)



English

Ukrainian

Chinese

~~Python~~

~~C++~~

Understanding
Generating

Natural language processing (NLP)

A dashed blue arrow originates from the word "Understanding" and points to the word "processing" in the phrase "Natural language processing (NLP)". A dashed blue oval encircles the word "processing".

Natural language processing (NLP)

NLP is not neuro-linguistic programming!

	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



Explain quantum computing in simple terms



Quantum computing is a type of computing that uses the principles of quantum mechanics to process information. Traditional computing uses bits, which can represent either a 0 or a 1. Quantum computing, on the other hand, uses quantum bits or qubits, which can represent a 0, a 1, or both at the same time.



This allows quantum computers to perform certain types of calculations much faster than traditional computers. For example, quantum computers can factor large numbers exponentially faster than classical computers, which is an important task in cryptography.

In addition to faster computations, quantum computing has the potential to solve certain problems that are difficult or impossible for classical computers to solve. However, building and programming a quantum computer is a very difficult task and requires a deep understanding of quantum mechanics.



Turn this explanation into a poem



Quantum computing, what a feat!
A new way to process and compute,
Using qubits, not just bits,
To speed up tasks that once were mute.



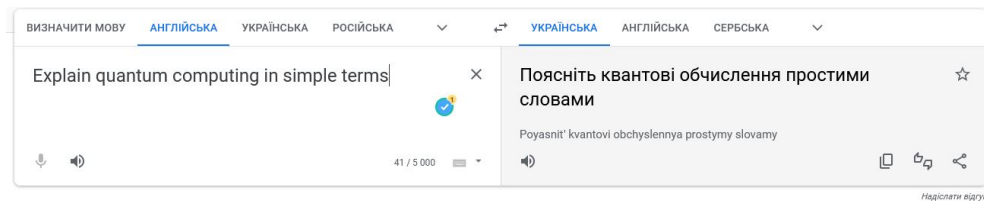
With bits, we had a simple way,
To represent just 0's and 1's.
But qubits add a third state,
Allowing for faster runs.

Quantum computers can do the work,
That once took ages to complete.
Factoring large numbers with ease,
And other tasks that are discrete.

But building such a wonder is hard,
And programming it is quite a feat.
With quantum mechanics as our guide,
We can make it all complete.

Large language models / chatbots

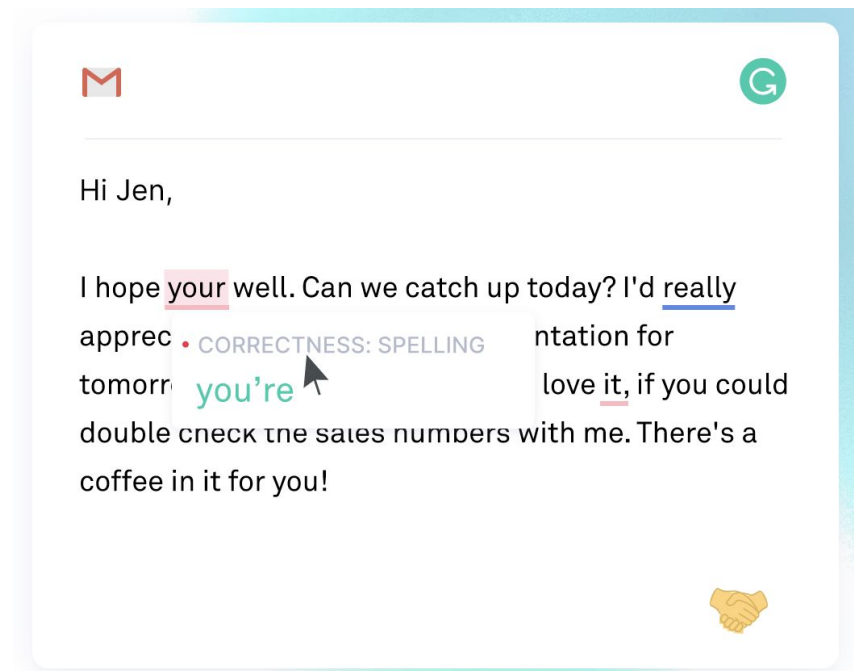
Machine translation



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

What would you like to do with ElevenLabs?

Select all that apply



Text to speech



Audiobooks



Music



Sound effects



Dubbing



Voice overs



Voice cloning



Speech to text

Code completion

SWE-Agents

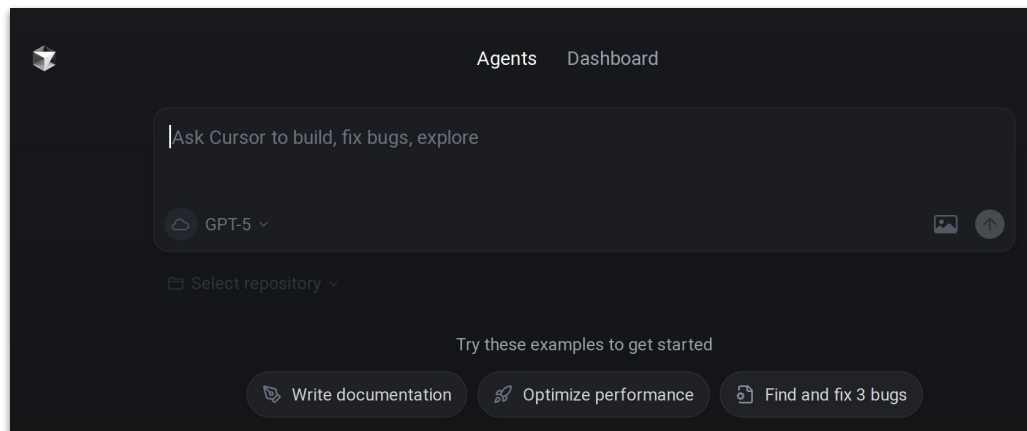
Code generation

Bug fixing

Code optimization

Documentation

...



Language is hard

Ambiguity

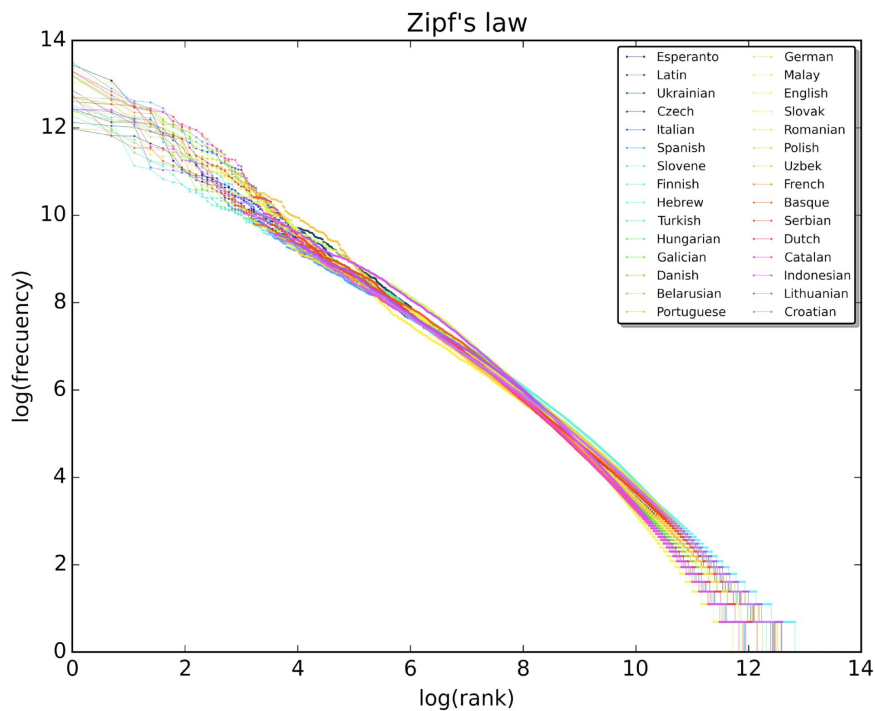
Я їм суп з куркою.

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

~50% have no written form

Sparsity

New dialects emerge

Neologisms

Over 7000 languages

Ambiguity

Sparsity

Neologisms

Over 7000 languages

Cultural references

Sarcasm, irony

...

Practical: Many applications

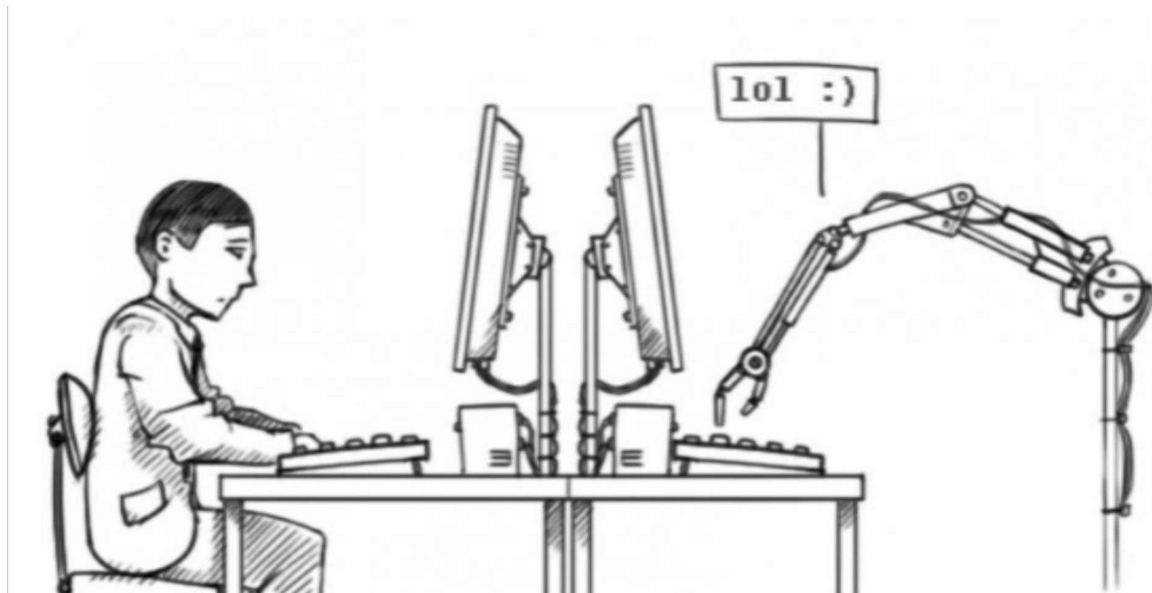
Research: Many open questions, path to AGI (maybe)

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

Exciting progress in recent years

NLP is fun!

Turing test



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

from existing texts

Winograd schemas

from dialogs

Acquiring new knowledge

Turing test

Winograd schemas

Acquiring new knowledge

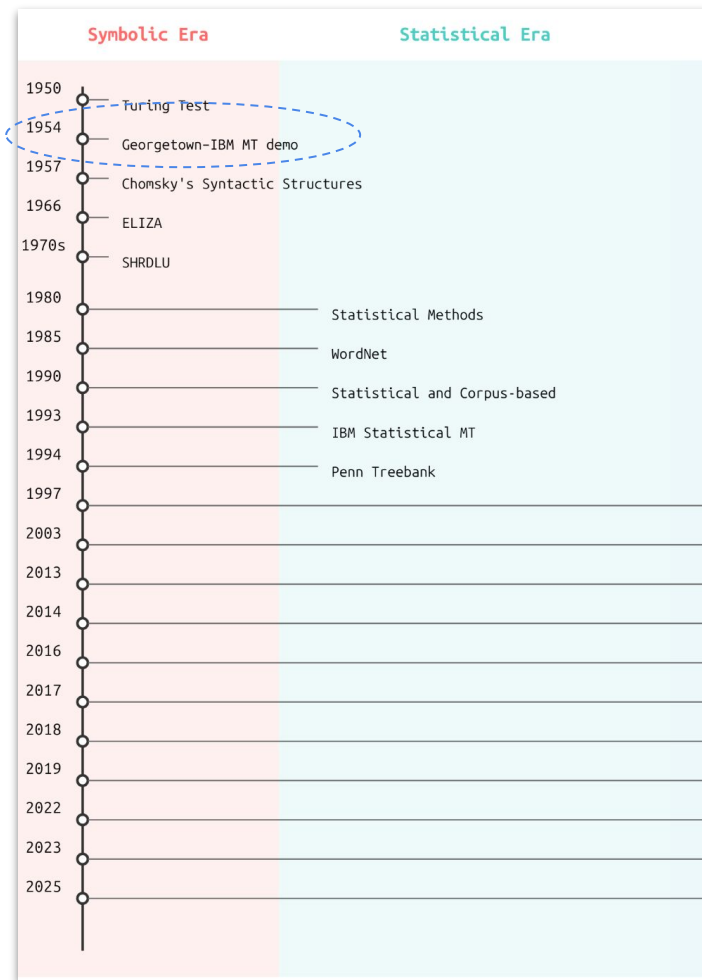
Some disagree 🙅

“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)

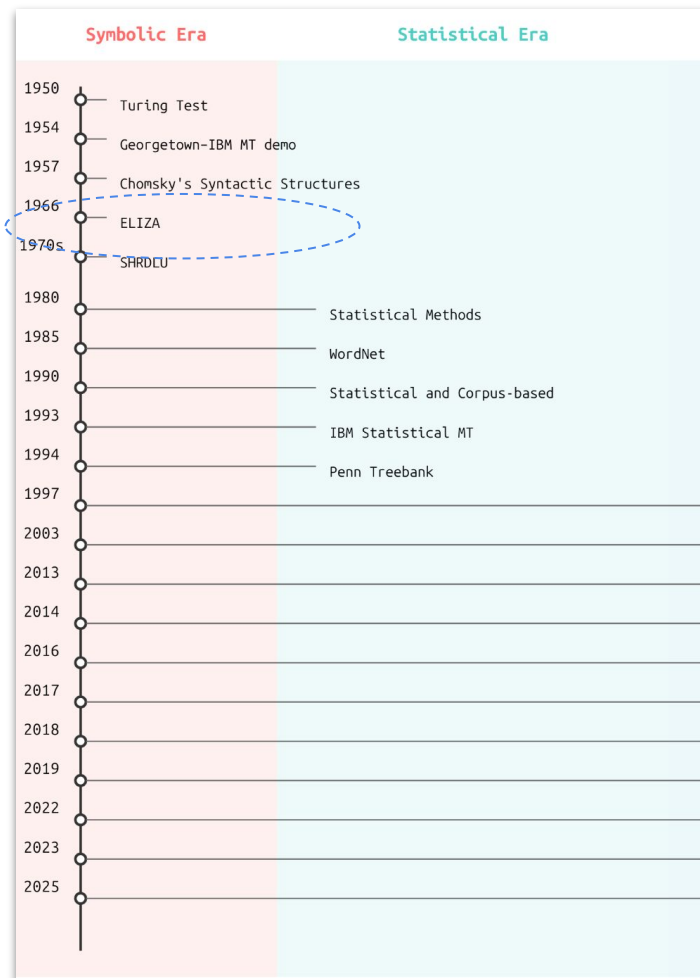
Punch cards for input

> 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



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?

>

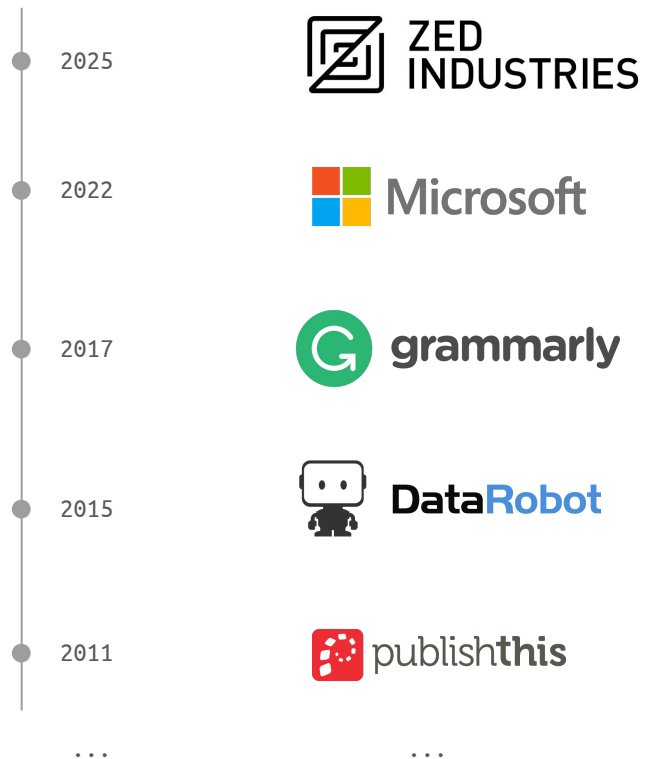
Logistics

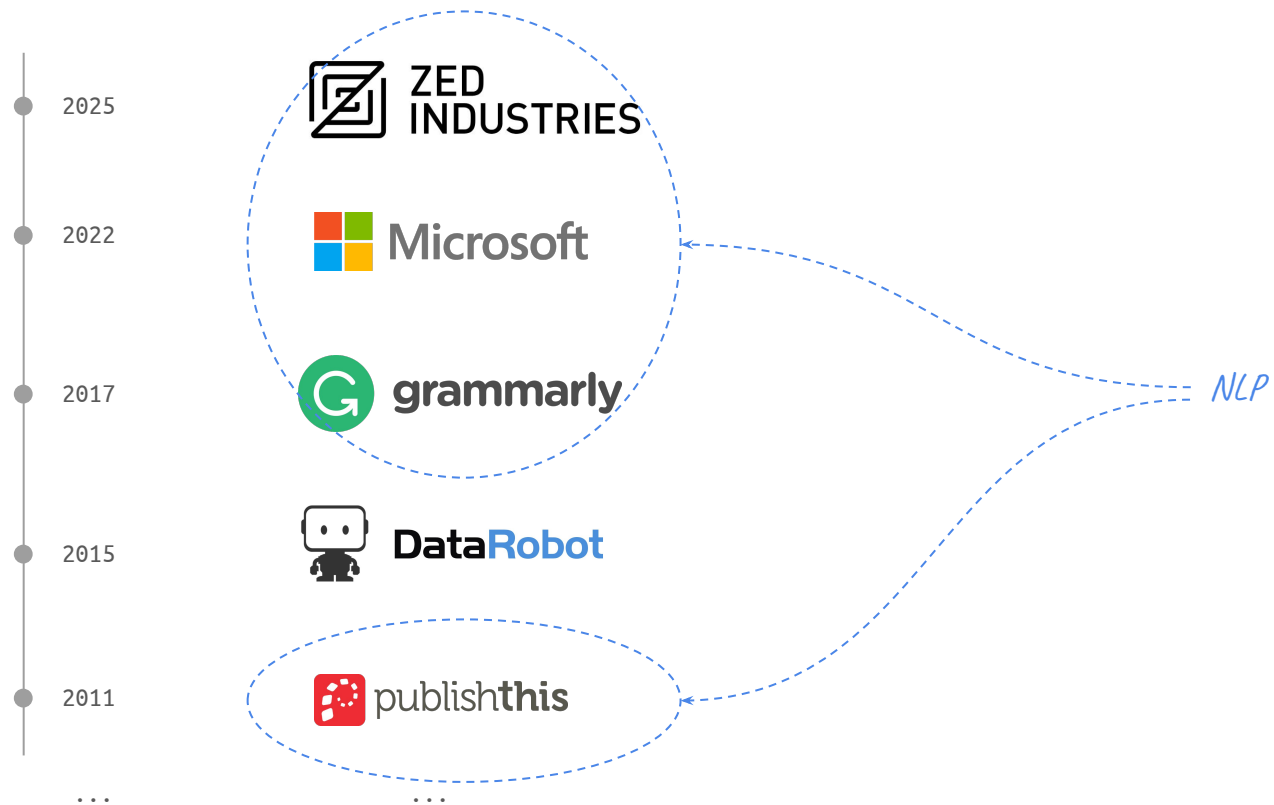
Oleksiy Syvokon

Сивоконь Олексій Олегович

AI Engineer @ Zed Industries

oleksii.o.syvokon@lpnu.ua





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

deep learning for NLP

10. Large language models (LLM)

-
11. Prompt engineering

12. Retrieval-augmented generation

13. Agents

LLMs

14. Preference optimization

15. Modern Transformer and alternative architectures

16. Multimodality

Rule-based NLP

Shallow ML NLP

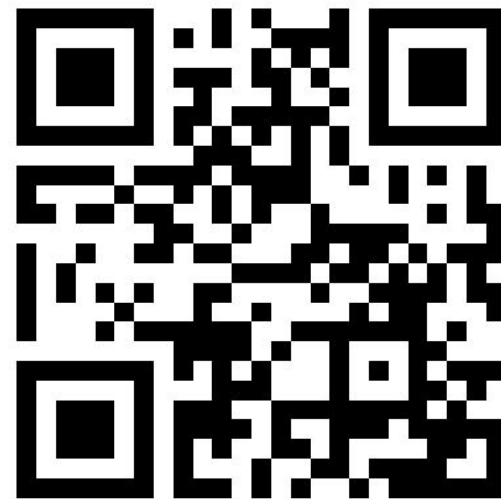
Linguistics and computational linguistics

What's next

Questions?

Join our Discord server

Text classification



<https://discord.gg/xXHnAry6>