Introduction to language and natural language processing (NLP)

Jasper Kyle Catapang
AI Engineer - Silicon Valley, USA
MA Applied Linguistics Student - University of Birmingham

August 19, 2021

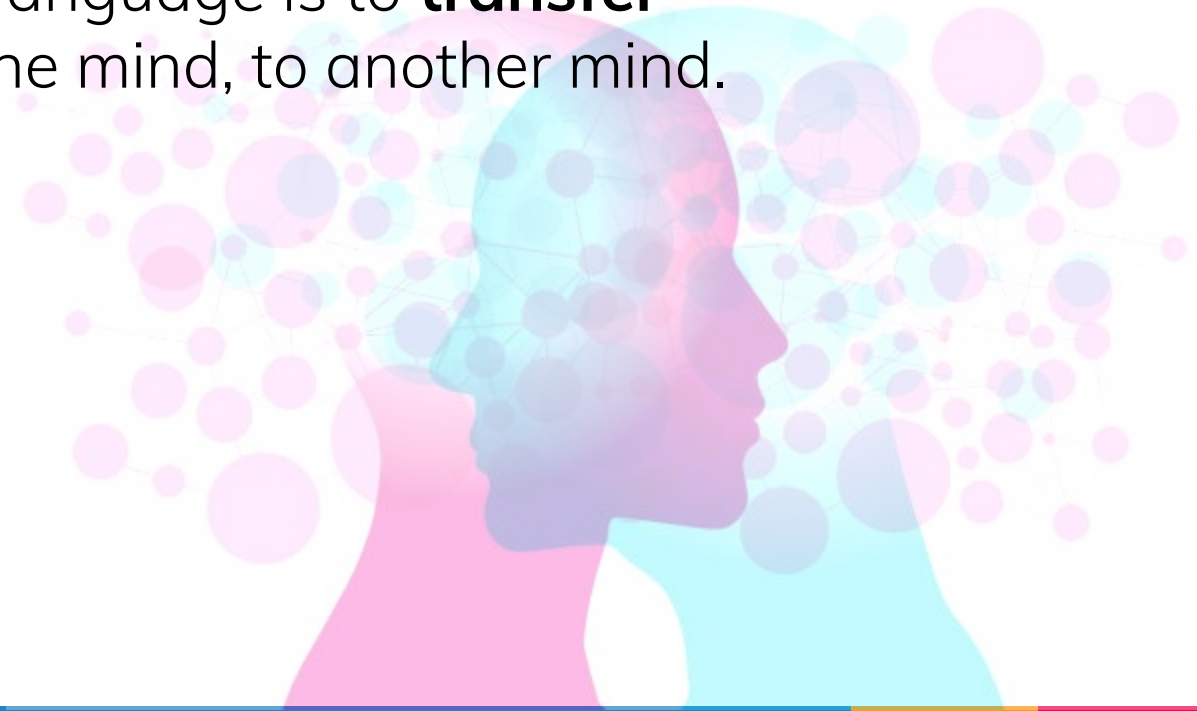
What is language?

“the principal method of human **communication**, consisting of **words** used in a **structured** and conventional way and conveyed by speech, writing, or gesture”

- Oxford Languages

What is language?

The main use of language is to **transfer thoughts** from one mind, to another mind.



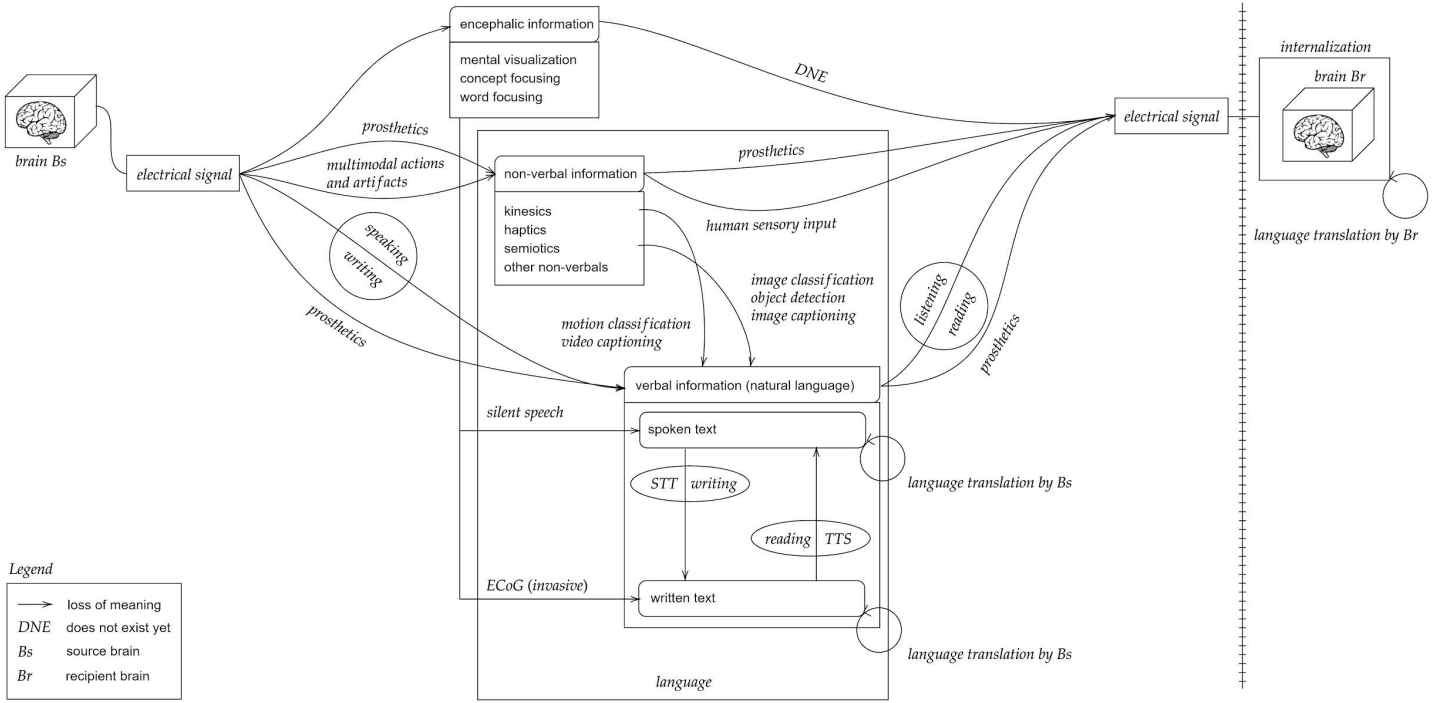
What is language?

Thought comes first, while language is an **expression**. There are certain **limitations** among language, and humans **cannot express all** that they think (Gleitman, 2005).

Propagation of meaning

Meaning is therefore **lost** with every medium or channel it goes through, from a source's mind to a recipient's mind.

Propagation of meaning model



Catapang, 2021

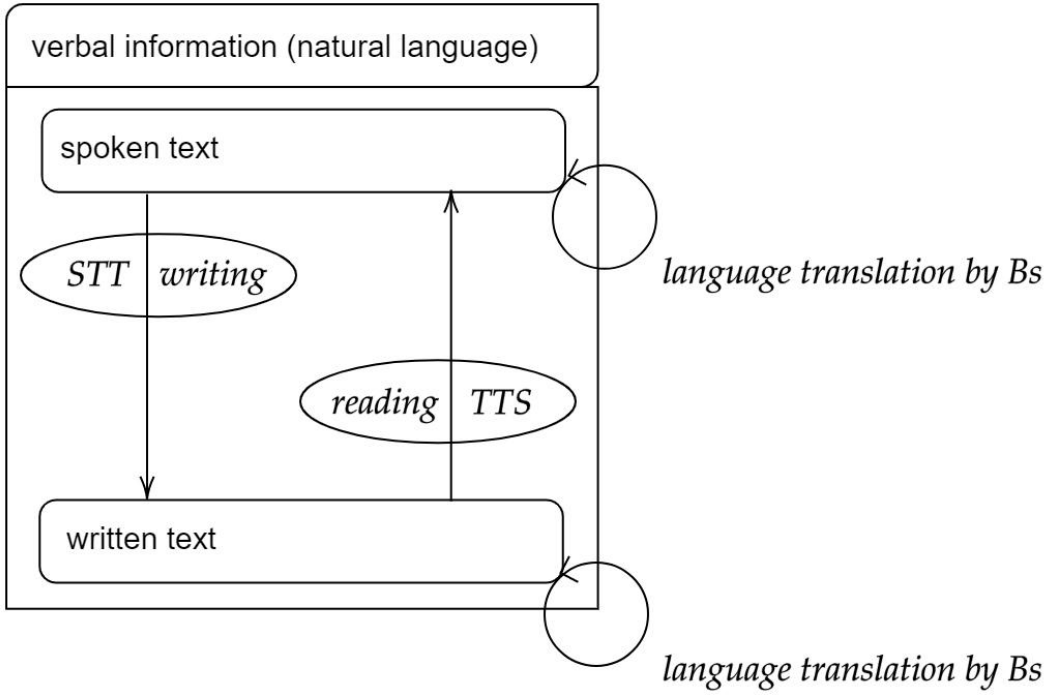
Non-verbal to verbal example

The K-pop girl group Twice is dancing to their new song.

A picture is worth a thousand words

https://openai.com/blog/dall-e/

Natural language

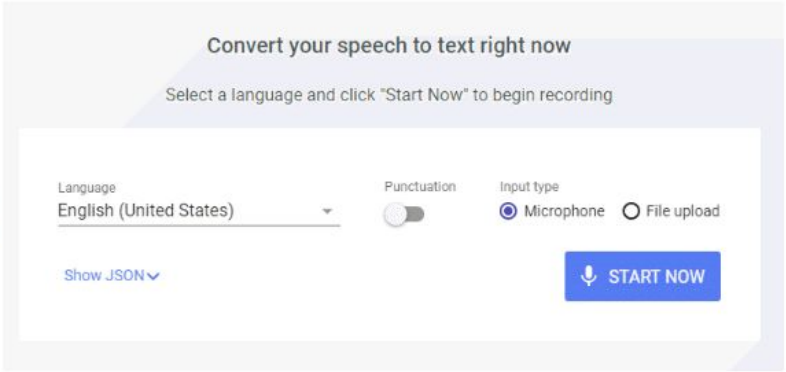


Catapang, 2021

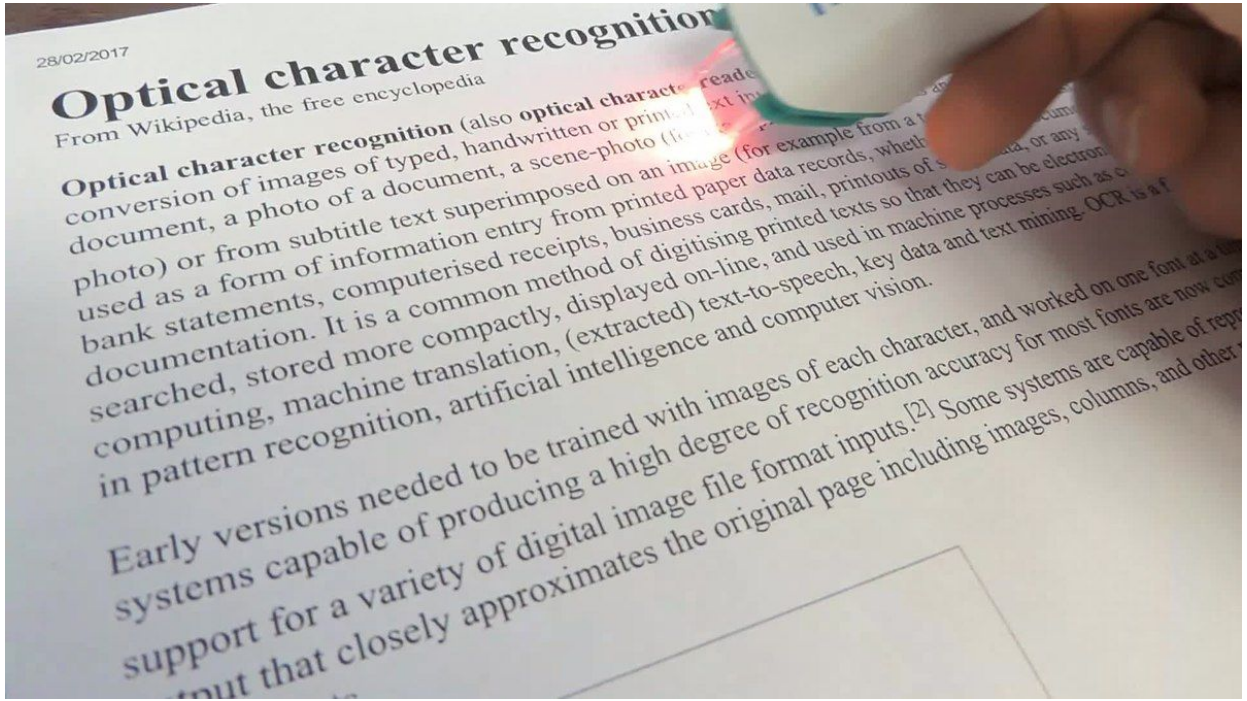
Speech to text

Google Cloud

Speech to Text



Optical character recognition (OCR)



Lost in translation

English: We bring you this show through the cooperation of our sponsors

Japanese: この番組はご覧のスポンサーの提供でお送りします

Korean: 이 프로그램은 보시는 스폰서 제공으로 보내드립니다

Filipino: Ang programang ito ay ipinadala ng sponsor na nakikita mo

Google Translate API

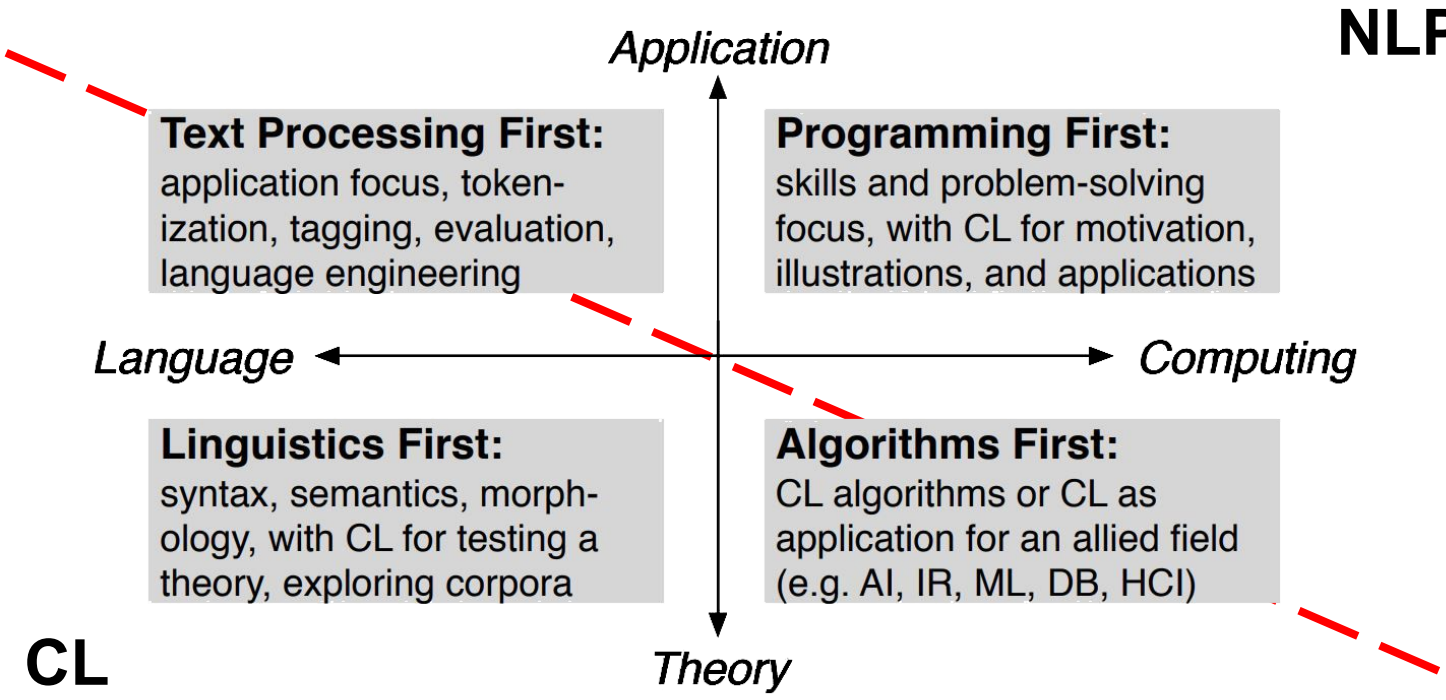
Natural language processing (NLP)

The goal of natural language processing is to get computers to perform useful tasks involving human language, tasks like **enabling human-machine** communication, **improving human-human** communication, or simply doing useful processing of text or speech (Jurafsky and Martin, 2009).

Additional goal of NLP

Keep the loss of meaning at a minimum to maximize communication.

NLP vs. computational linguistics (CL)



Bird, 2008

Why learn NLP?

- Many applications!
- Major contributor to growing AI market
- Demand is high
- Understanding how machines can learn language more efficiently and correctly

Why learn NLP?

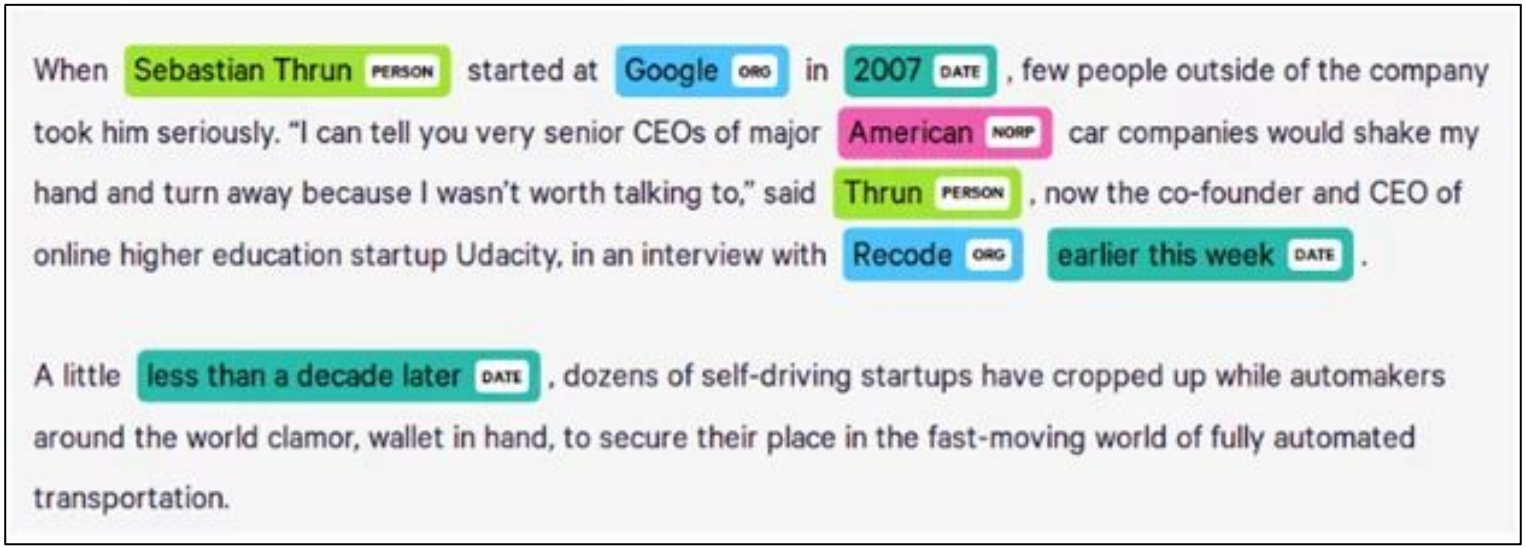
- Many applications!
- Major contributor to growing AI market
- Demand is high
- Understanding how machines can learn language more efficiently and correctly

Applications

- Named entity recognition
- Machine translation
- Question answering
- Chatbots
- Summarization

And many more!

Named entity recognition



Applications

- Named entity recognition
- Machine translation
- Question answering
- Chatbots
- Summarization

And many more!

Why learn NLP?

- Many applications!
- Major contributor to growing AI market
- Demand is high
- Understanding how machines can learn language more efficiently and correctly

Growing AI market

The global artificial intelligence (AI) software market is forecast to grow rapidly in the coming years, reaching around **126 billion** U.S. dollars by 2025. The overall AI market includes a wide array of applications such as natural language processing, robotic process automation, and machine learning (Liu, 2020).

Why learn NLP?

- Many applications!
- Major contributor to growing AI market
- Demand is high
- Understanding how machines can learn language more efficiently and correctly

Demand is high

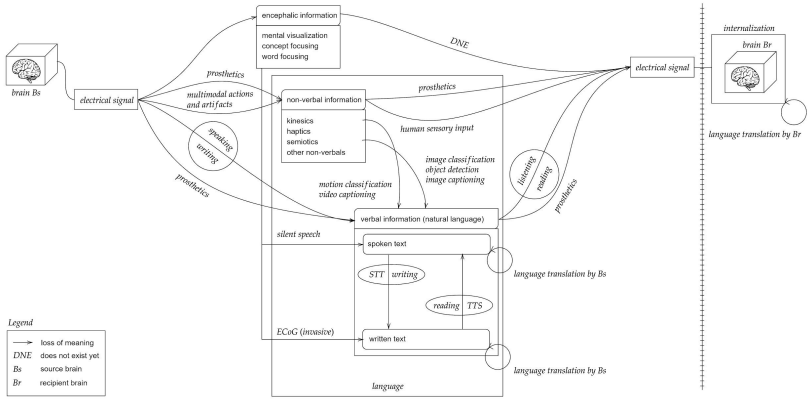
- The pandemic catalyzed digitalization of all companies and businesses
- This includes the digitalization and processing of text data
- Jobstreet openings with “natural language processing” as query, pay PHP 50,000 at the least (as of August 12, 2021).

Why learn NLP?

- Many applications!
- Major contributor to growing AI market
- Demand is high
- Understanding how machines can learn language more efficiently and correctly
  - and by extension... customers also.

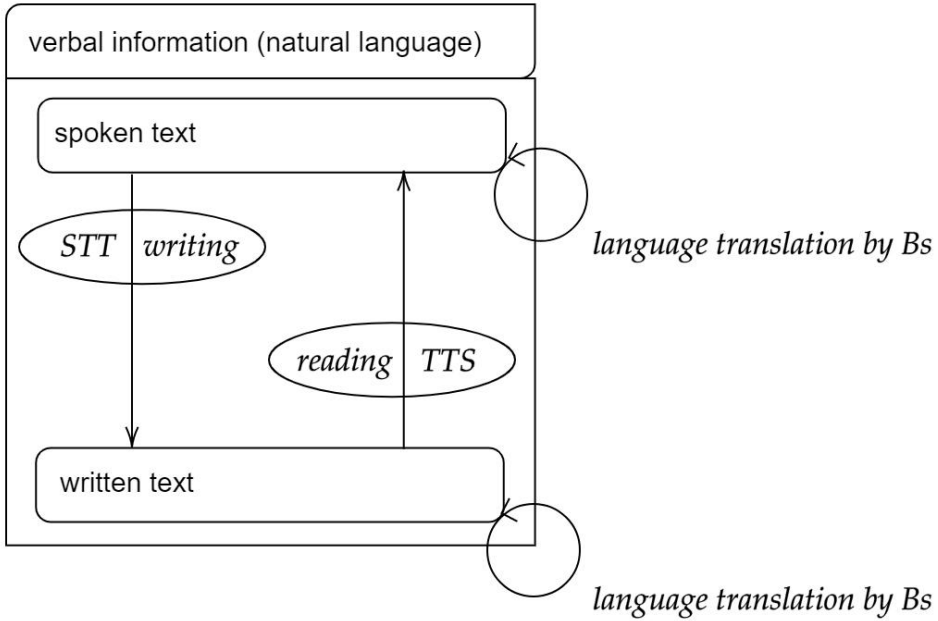
Why learn NLP?

- Many applications!
- Major contributor to growing AI market
- Demand is high
- Understanding how machines can learn language more efficiently and correctly
- Promote the diversity and preservation of Philippine languages



Introduction to language and natural language processing (NLP)

Building blocks of natural language



- characters
- morphemes
- words

To be discussed by the next set of slides on: **Words, morphology, and the lexicon**

References

- Bird, S. 2008. *Defining a core body of knowledge for the introductory computational linguistics curriculum*. In Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics (TeachCL '08). Association for Computational Linguistics, USA, 27–35.
- Gleitman, L. 2005. *Language and thought*. Cambridge Handbook of Thinking and Reasoning.
- Jurafsky, D., and Martin, J. H. 2009. *Speech and Language Processing, 2nd edition*. Stanford University and University of Colorado at Boulder.
- Liu, S. 2020. *AI market size 2018-2025*. Retrieved from: https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/