

## Introduction

PA154 Language Modeling (1.1)

**Pavel Rychlý**

pary@fi.muni.cz

February 16, 2023

## PA154 – Technical Informations

- Slides in IS  
<https://is.muni.cz/auth/el/fi/jaro2023/PA154/>
- Final written exam (online)  
50 points, 25 points for E
- optional individual projects  
up to 25 points

Pavel Rychlý • Introduction • February 16, 2023

2 / 9

## Individual projects

- presentation on a new research in language modeling
- small project as a part of bigger collaborative projects
  - neural machine translation
  - lexical acquisition
- small task
  - describe errors in ChatGPT
  - annotation of a language resource

Pavel Rychlý • Introduction • February 16, 2023

3 / 9

## Language model

- model
  - (mathematical) abstractions
  - similar/same behavior of modeled object
- language model
  - model a natural language

Pavel Rychlý • Introduction • February 16, 2023

4 / 9

## Language models—what are they good for?

- assigning scores to sequences of words
  - predicting words
  - generating text
- ⇒
- statistical machine translation
  - automatic speech recognition
  - optical character recognition

Pavel Rychlý • Introduction • February 16, 2023

5 / 9

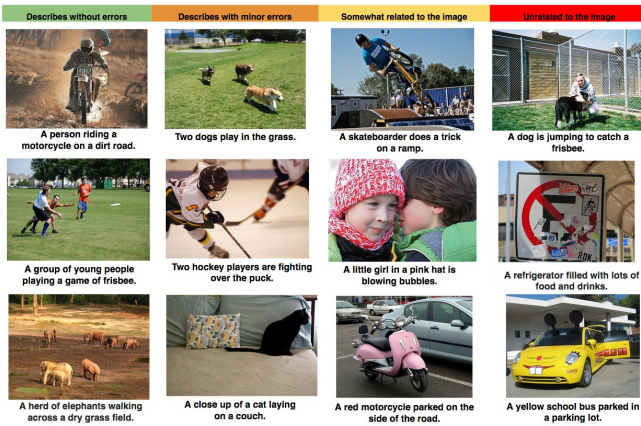
## Predicting words

Do you speak ...  
 Would you be so ...  
 Statistical machine ...  
 Faculty of Informatics, Masaryk ...  
 WWII has ended in ...  
 In the town where I was ...  
 Lord of the ...

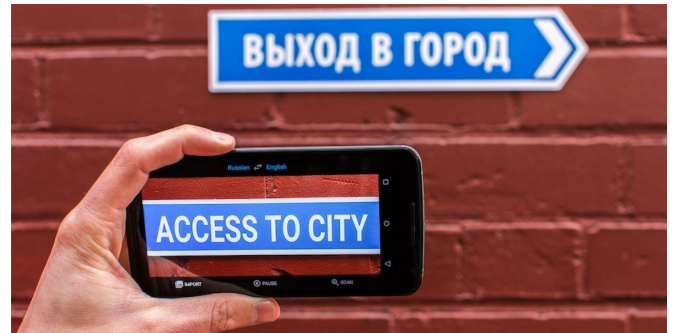
Pavel Rychlý • Introduction • February 16, 2023

6 / 9

## Generating text



## MT + OCR



## Language models – probability of a sentence

- LM is a probability distribution over all possible word sequences.
- What is the probability of utterance of  $s$ ?

### Probability of sentence

$p_{LM}(\text{Catalonia President urges protests})$

$p_{LM}(\text{President Catalonia urges protests})$

$p_{LM}(\text{urges Catalonia protests President})$

...

Ideally, the probability should strongly correlate with fluency and intelligibility of a word sequence.