

# Machine Translation

PV061

**Pavel Rychlý**

NLP Centre, FI MU

20 Sep 2023

# Optional Assignments

You can earn extra points for the exam.

# Errors in translations

(1-2 points for a group)

- find errors in machine translation
- choose any system (Google, DeepL, ...)
- translate a document
- report and describe errors
- try to find shortest sentence containing that error

# Errors in training data

(1-2 points for a group)

- look at some training data for translation
- find errors
  - wrong alignments
  - wrong language
  - non-fluent sentences

# Compare small optimizations

(max 15 points)

- find a NMT system
- train with original setup
- train with a small optimizer
  - different tokenizer (BPE/SentencePiece/HFT)
  - shared/separate embeddings
- compare results

# Replicate SotA results on own data

(max 15 points)

- find a NMT implementation
- run on own data (Cz-En ficet)
- compare results

**MUNI**

FACULTY

OF INFORMATICS