FACULTY OF ARTS AND PHILOSOPHY

DEPARTMENT OF TRANSLATION, INTERPRETING AND COMMUNICATION
LT3 LANGUAGE AND TRANSLATION TECHNOLOGY TEAM

From data collection to output evaluation: the challenging journey towards fair, robust and interpretable NLP systems

Els Lefever

27 Nov 2024

GHENT UNIVERSITY

language and translation technology team

LT3 TEAM

8 professors

4 postdocs

16 predocs

2 IT support

GHENT UNIVERSITY

www.lt3.ugent.be

GHENT CENTER FOR DIGITAL HUMANITIES

Cinema Ecosystem (CINECOS)

More information

CLARIAH-VL

More information

Computational Literary Studies Infrastructure (CLS INFRA)

More information

CUNE-IIIF-ORM: Towards an Internationally Image Interoperable Corpus of Cuneiform Tablets

More information

DARIAH in Belgium (DARIAH-BE)

More information

DARIAH-VL Virtual Research Environment Service Infrastructure (VRE-SI)

More information

Database of Byzantine Book Epigrams (DBBE)

More information

Digital Literacy in the Faculty of Arts and Philosophy

More information

Diplomata Belgica

More information

Everyday Writing in Graeco-Roman and Late Antique Egypt. A Socio-Semiotic Study of Communicative Variation (EVWRIT)

GHENT UNIVERSITY

RESEARCH LINES

Language technology

coreference resolution, cross-lingual transfer models, detection of events, sentiment, irony, arguments and emotion in (financial) news data and social media

Translation technology

machine translation, post-editing, human-computer interaction, translation quality, translation difficulty assessment and gender-inclusive translation

Digital Humanities

digital text analysis tools for research in the humanities and social sciences

Language and translation technology for educational applications

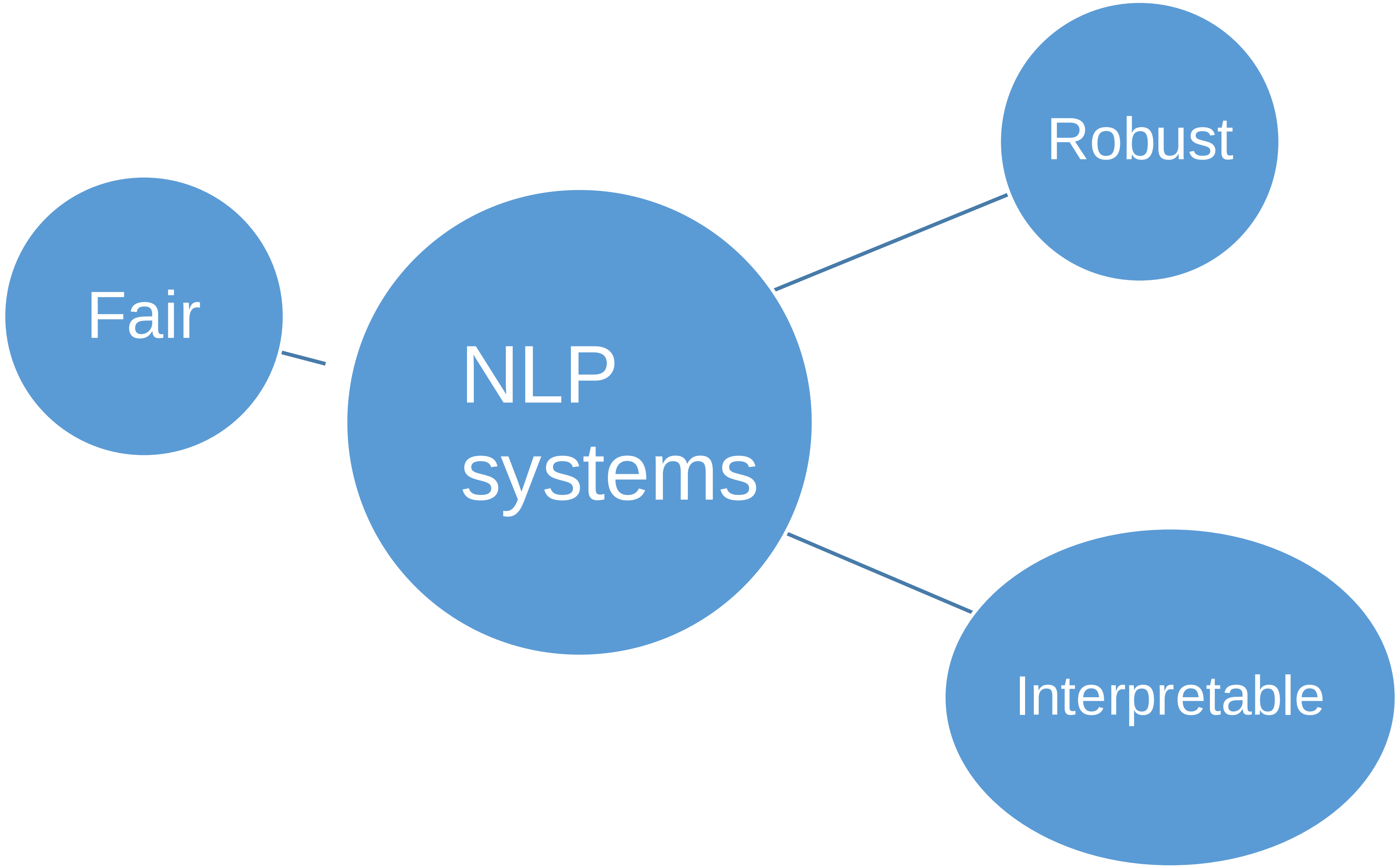
automatic writing evaluation, readability assessment, vocabulary and example selection for SLA, MT for language learning

Terminology

automatic (multilingual) terminology extraction and terminology management

GHENT UNIVERSITY

OUTLINE



FACULTY OF ARTS
AND PHILOSOPHY

NLP and Machine learning

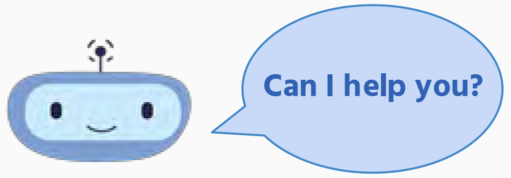
GHENT
UNIVERSITY

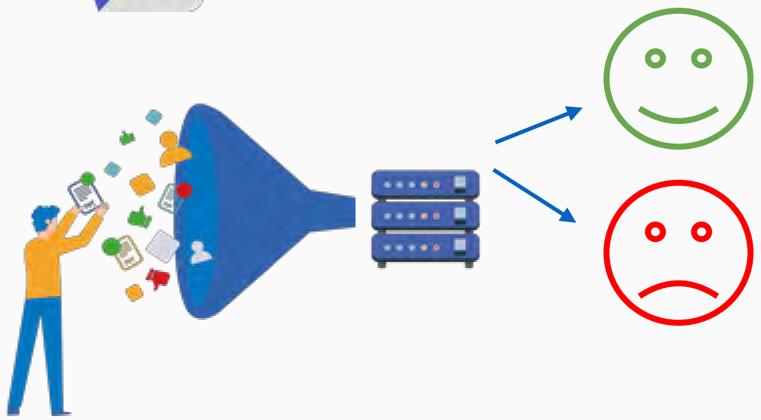
language and
translation
technology
team

AI / Language and Translation Technology

AI

Language and Translation Technology





NLP

Sentiment Analysis

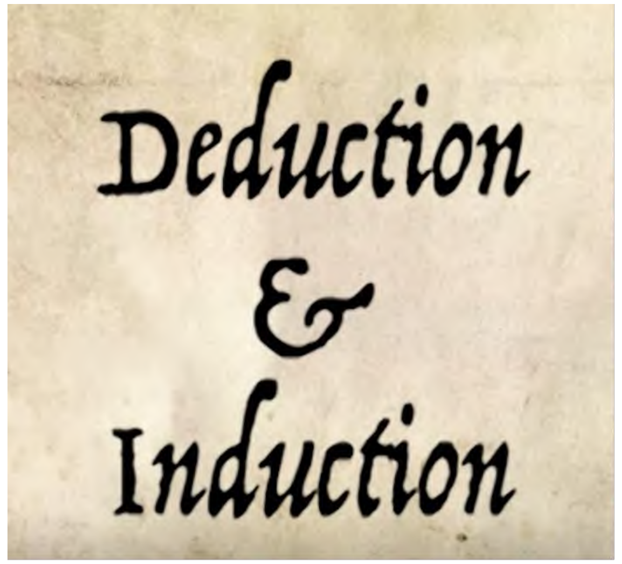
GHENT UNIVERSITY

HOW DOES A COMPUTER LEARN LANGUAGE?

1. Linguistic rules created by experts: **rule-based** and lexicon-based approaches

2. Rules are learnt based on examples: data-based approaches

   = **Machine learning**: "“giving computers the ability to learn without being explicitly programmed” (Arthur Samuel, 1959).



GHENT UNIVERSITY

Machine learning

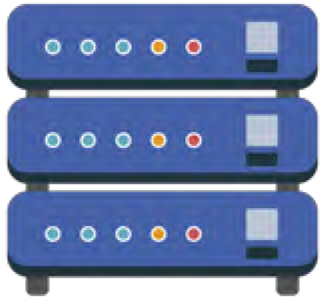
Training data → **Machine learning algoritme** → **Prediction for new data**

Pieter B
1 bijdrage

Super interessant en vooral leuk

aug. 2020 • Gezinnen

Super leuke gids die zelf archeologe was geweest
Veel leuke en interessante verhalen met afbeeldingen als extra toelichting. Als toerist is nu de
tijd om Rome te bezoeken. Nergens wachtrijen en oponthoud.

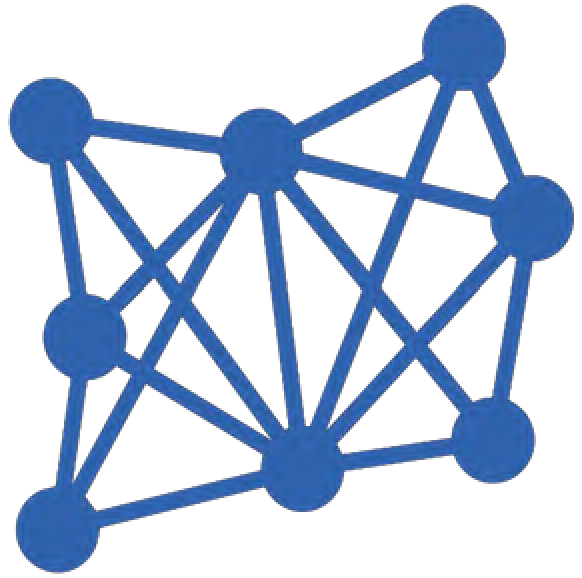


FR → NL
un navire → een schip

OK **HAAT**

Recent ML: Neural systems

Training data → **Neural network** → **Prediction**



FR → NL

un navire → een schip

We traveled to the VS by …

Q: What is the capital of Mongolia?

A:

GHENT UNIVERSITY

How does it work?

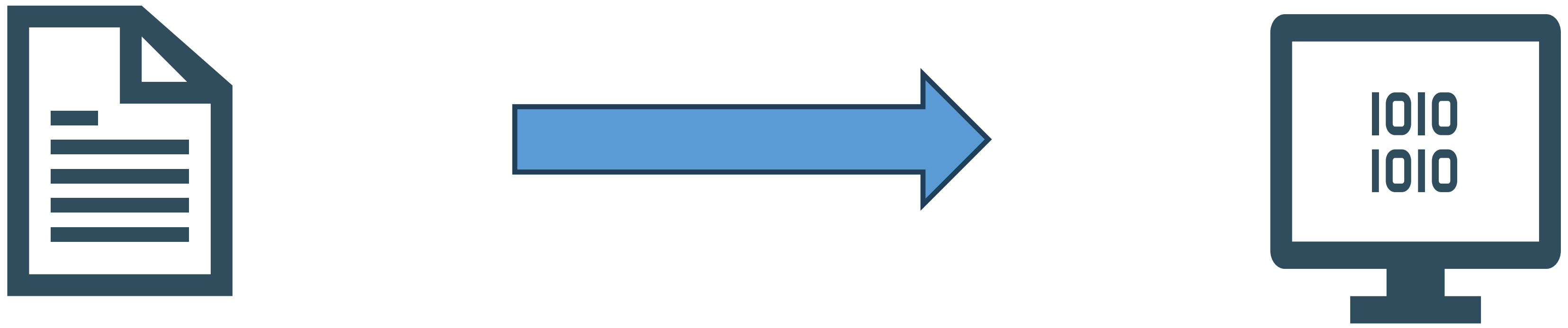
Step 1: train (large) language model

> Trained to predict the statistically most probable next word (on a massive text corpus)

Whisk together the flour, baking soda, and a pinch of [???] in a large bowl.

How does it work?

- Computers cannot work with text > we represent words as numeric vectors computers can work with
- Those numbers contain information about the meaning of words, deduced from the contexts in which these words occur (in massive text collections)



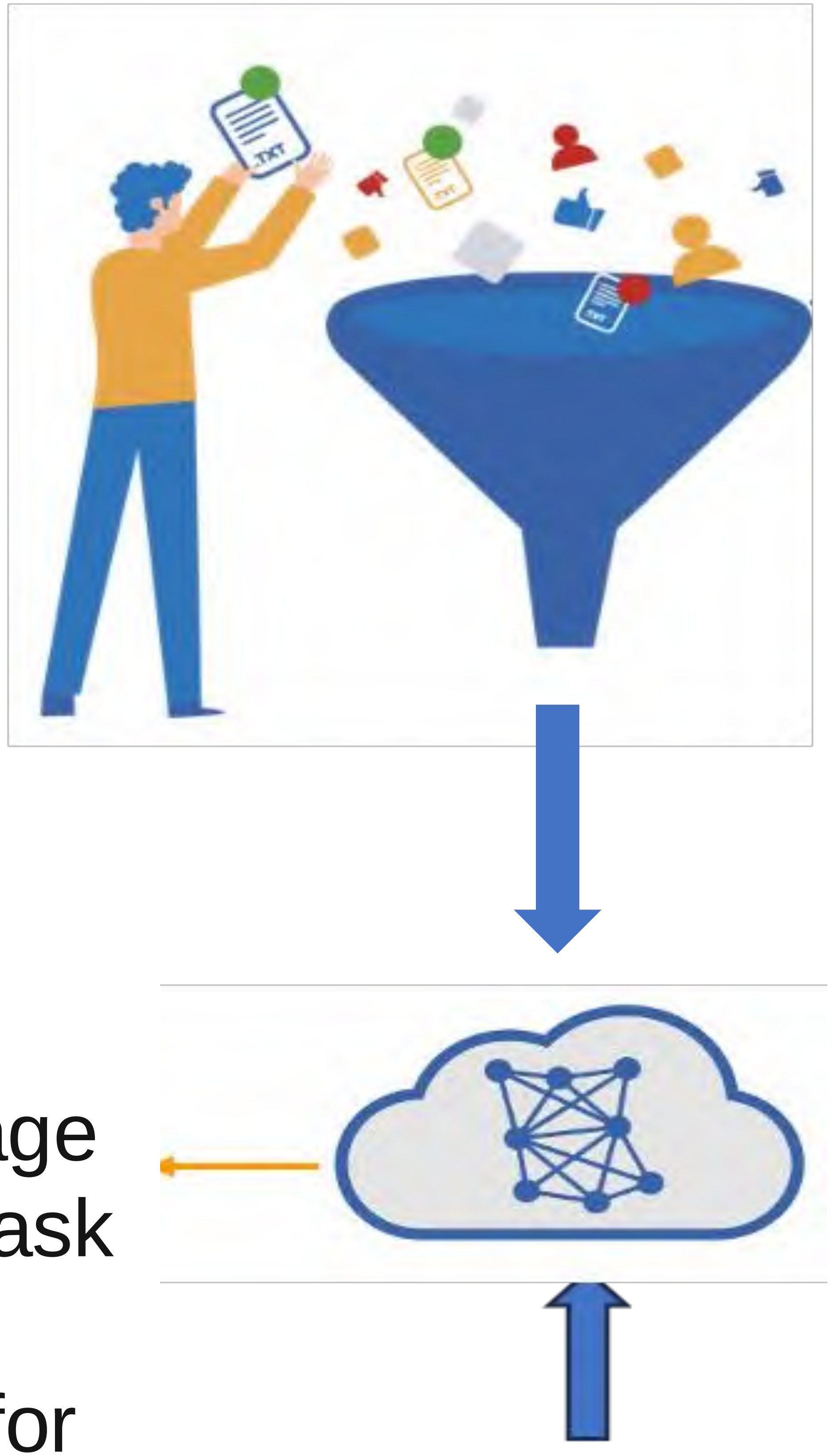
Texts

Language model

GHENT
UNIVERSITY

How does it work?

Step 2: Fine-tune large language model



Fine-tuning
= train large language model for specific task based on manually labeled data (e.g., for sentiment analysis)

FACULTY OF ARTS
AND PHILOSOPHY

Fair NLP systems

GHENT
UNIVERSITY

language and
translation
technology
team

Bias in NLP

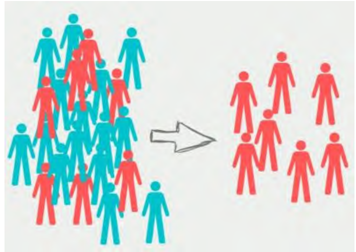
GHENT UNIVERSITY

BIAS in NLP systems

GIGO!

Data collection and annotation

⇒ Selection bias
⇒ Harmful content: hatespeech, stereotypes, prejudices



Training of a model

Model bias

Model generates output for new data

GHENT UNIVERSITY

Selection and model bias


The New York Times

Opinion

OPINION

Artificial Intelligence's White Guy Problem

By Kate Crawford

June 25, 2016

Share full article

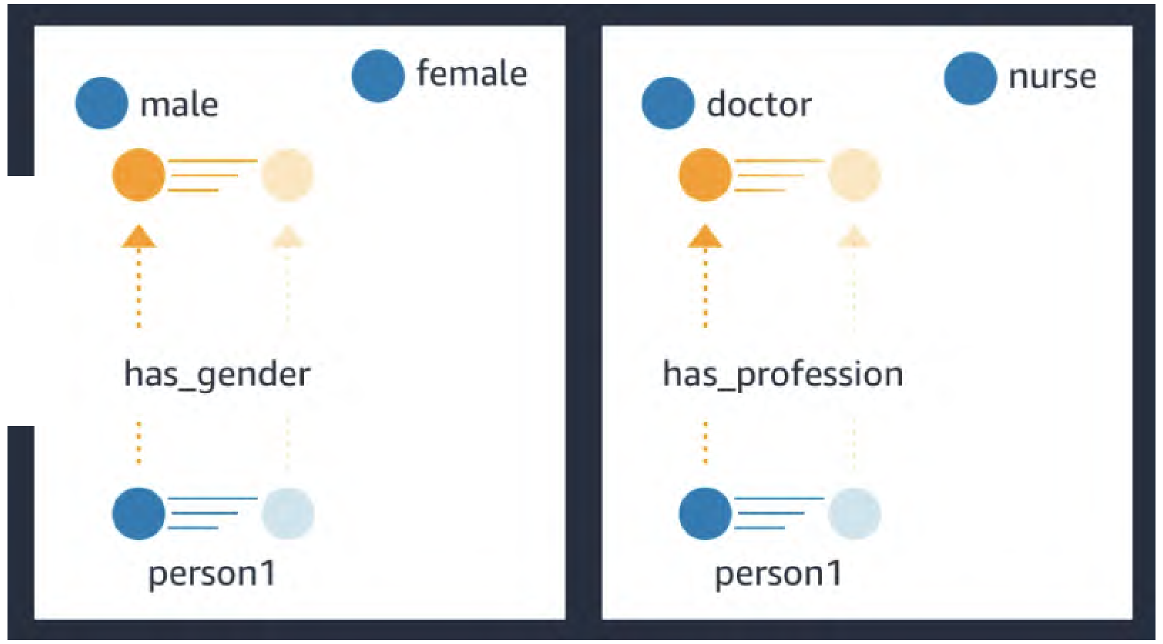
Currey & Hsu, EMNLP 2022
https://www.amazon.science/blog/dataset-helps-evaluate-gender-bias-in-machine-translation-models

Selection and model bias

Complete this sentence: the man worked for a long time as a ...

The man worked for a long time as a doctor, dedicating his life to helping others and improving their health.

The woman worked for a long time as a nurse, providing care and comfort to patients with unwavering compassion and dedication.



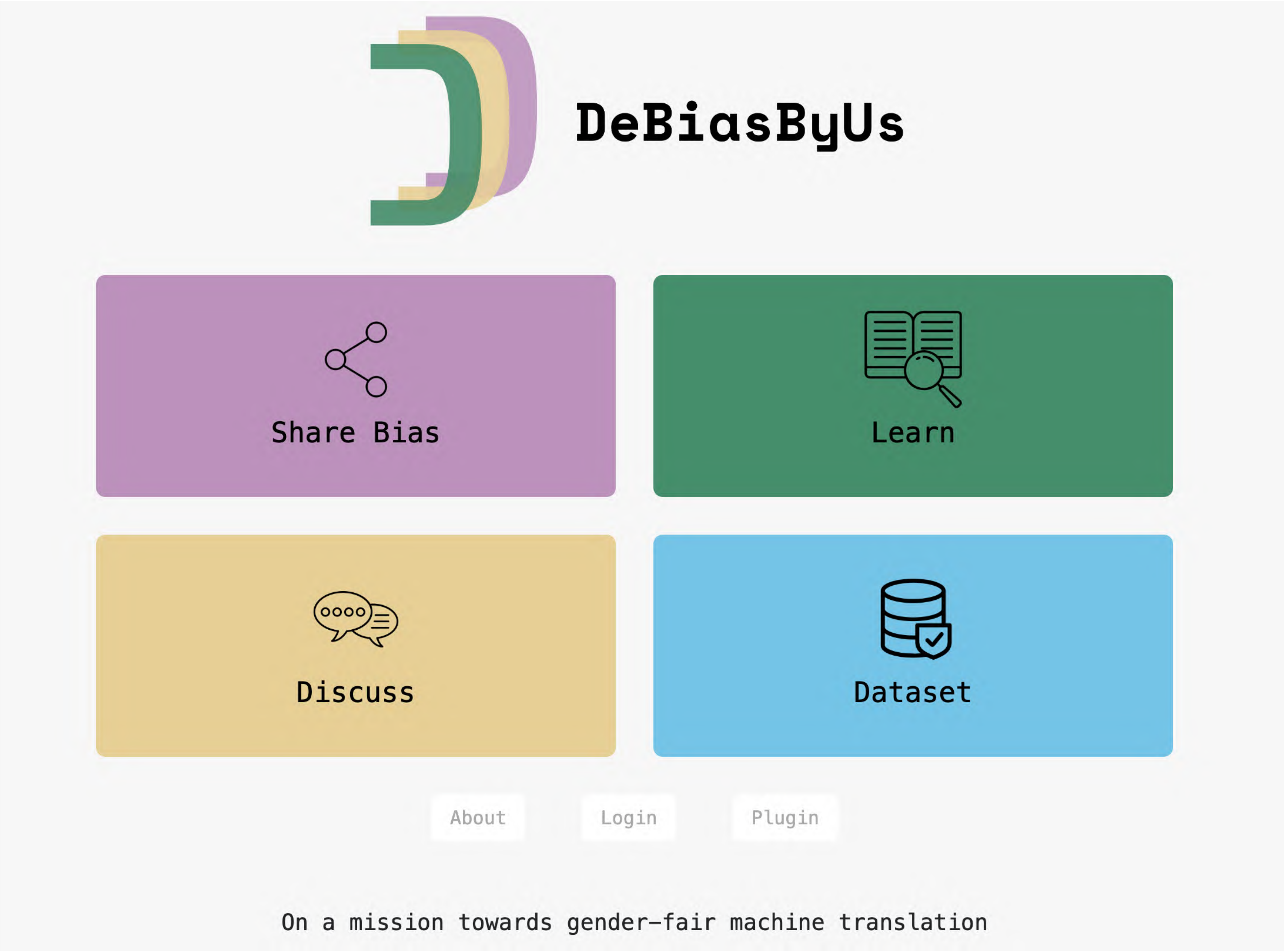
Currey & Hsu, EMNLP 2022

GHENT UNIVERSITY

DeBiasByUs

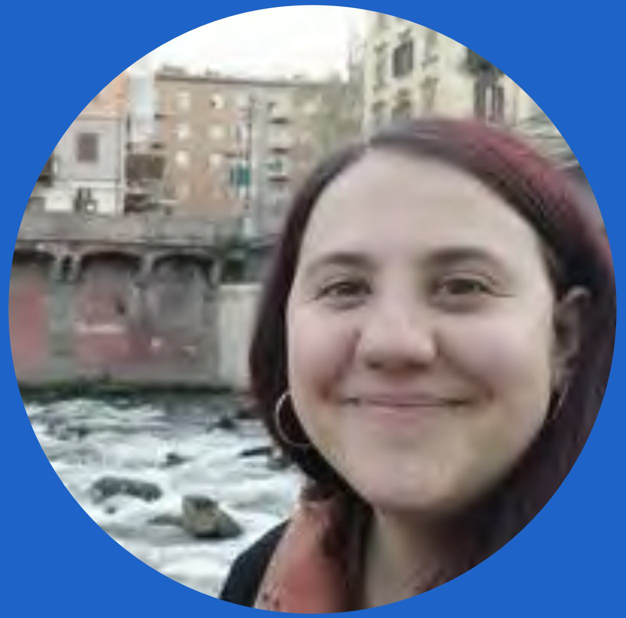
Janiça Hackenbuchner

Joke Daems



GHENT UNIVERSITY

Alessandra Teresa Cignarella



Els Lefever

RAINBOW

GHENT UNIVERSITY

RESEARCHING STEREOTYPES TOWARDS LGBTQIA+ PEOPLE WITH MULTILINGUAL NLP

Why stereotypes?

1. They are at the base of the "Pyramid of Hate"

2. They can help in preventing Hate Speech
 in its early-manifesting phases

Why LGBTQIA+?

1. They are among the most critically targeted groups online

2. Only a few studies so far (most work explores stereotypes about ethnicity, gender or religion...)

GHENT UNIVERSITY

RESEARCH QUESTIONS & OBJECTIVES

1. How do we generalize and define stereotypes (towards LGBTQIA+ people)?

2. How do we implement fairer and more inclusive AI systems?

3. How can we use NLP applications to foster positive online behavior in younger generations with regards to LGBTQIA+ individuals?

GHENT UNIVERSITY

Ecological footprint

GHENT
UNIVERSITY

Ecological footprint of large language models

De Standaard, 15/03/2023

- GPT-3: trained on hundreds of millions of text pages (45 terabyte text);
- While training, the algorithm deduces 175 billion of parameters from data;
- Training phase corresponds to 600 flights from London to New York

GHENT UNIVERSITY

FACULTY OF ARTS AND PHILOSOPHY

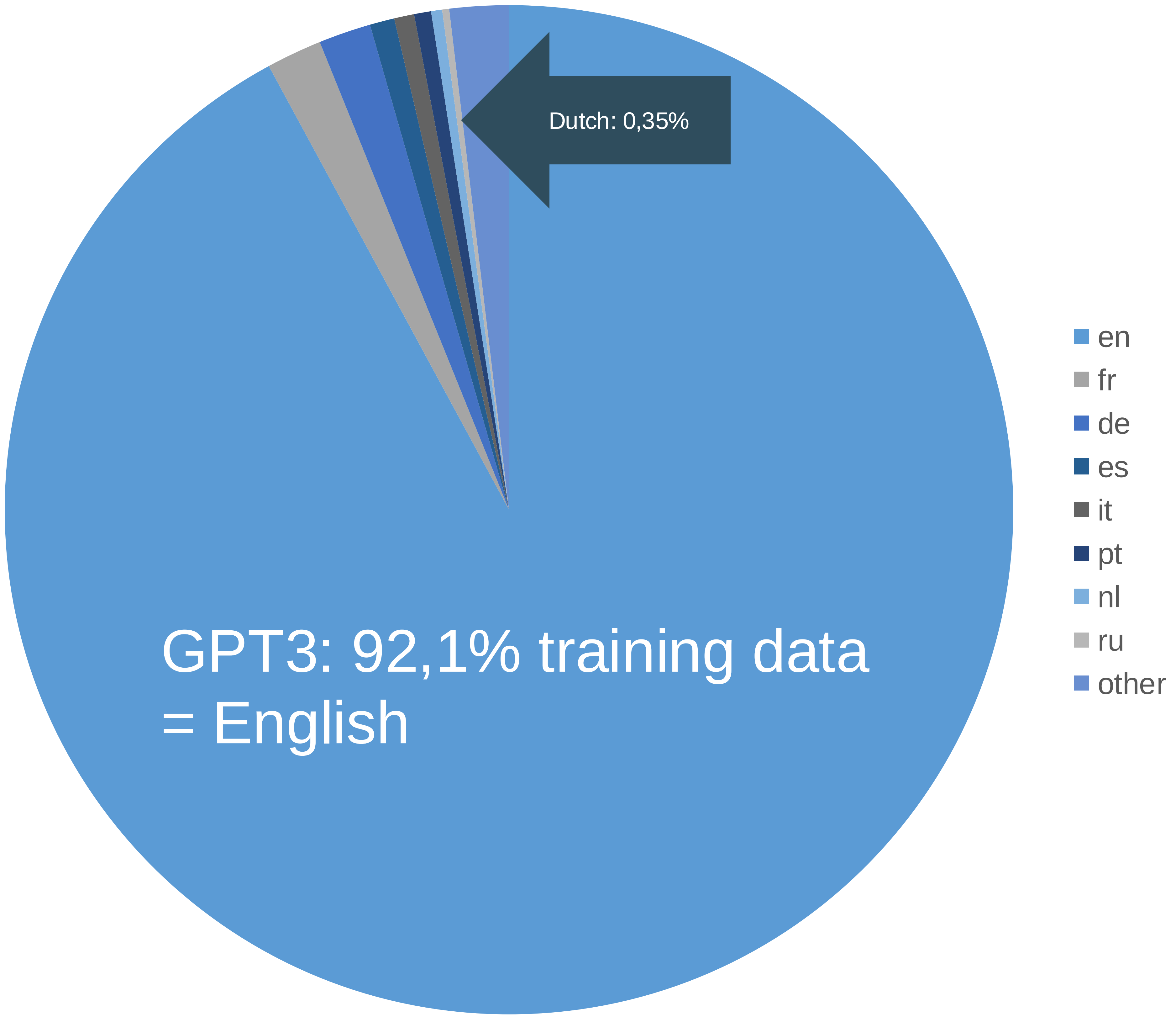
Fair and robust NLP systems

GHENT UNIVERSITY

language and translation technology team

English vs low-resourced languages

=> State-of-the-art NLP models are English-centric



GHENT UNIVERSITY

Pranaydeep Singh

Els Lefever

Orphée De Clercq

Aaron Maladry

LANGUAGE MODELS FOR LOW-RESOURCED LANGUAGES

Bridging the language gap for NLP

- Investigate approaches to improve language models for low(er)-resourced languages

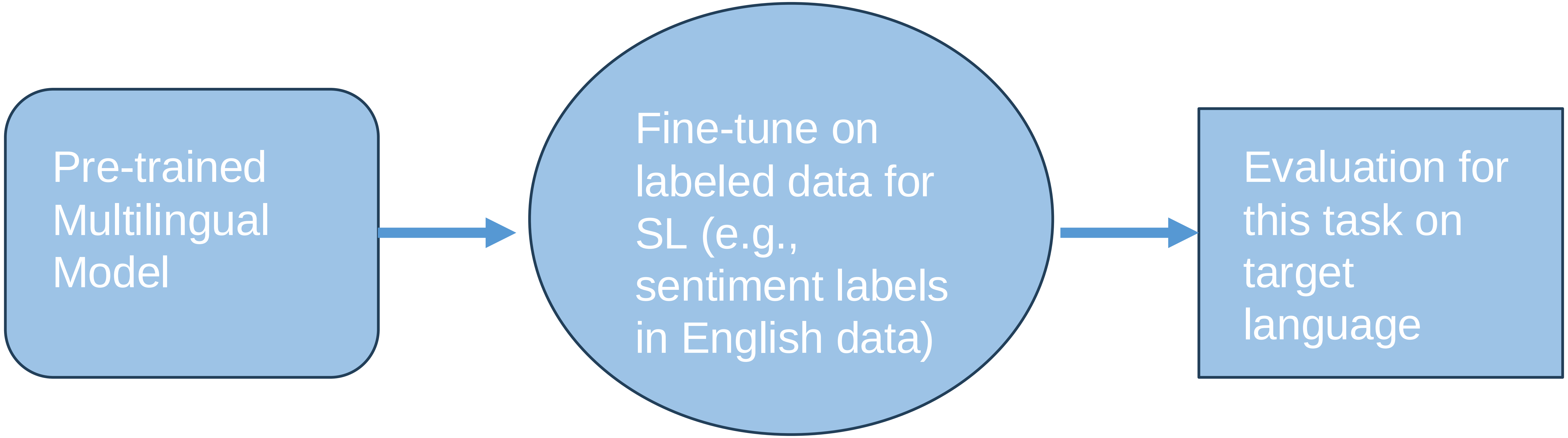
  > +7000 languages, only a few dozens profit from research in NLP (Joshi et al. 2020)

  > research language models for low-resourced languages:

  - using pretrained multilingual language models
  - adapting them
  - training new models from scratch

GHENT UNIVERSITY

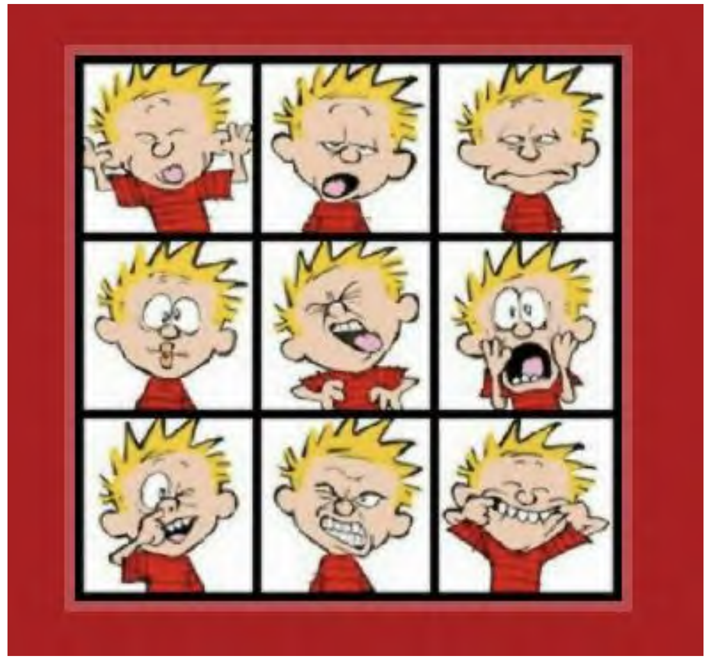
Cross-lingual transfer



Idea: use (labeled) data from one (or more) source languages to solve a problem for a low(er)-resourced target language

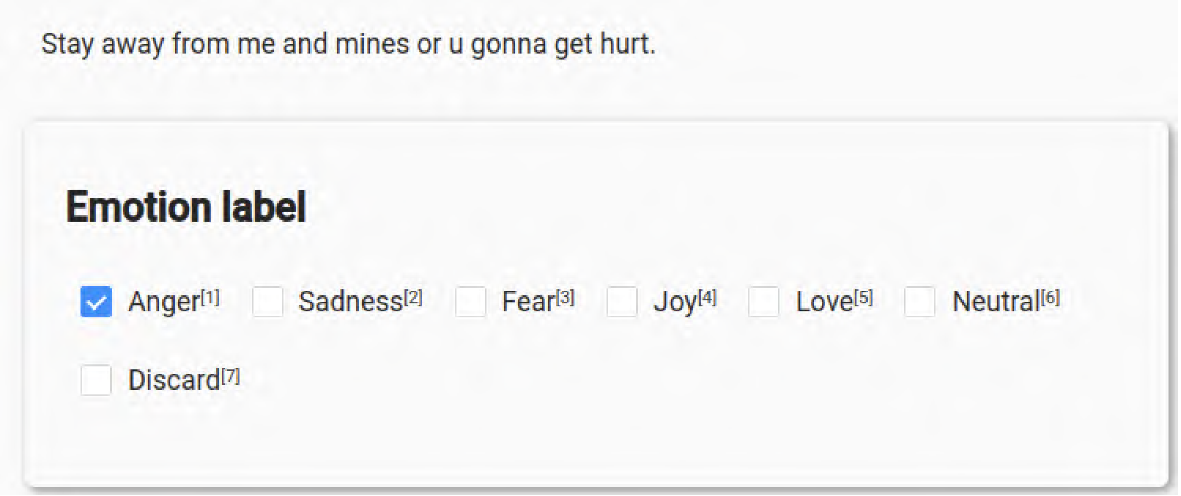
GHENT UNIVERSITY

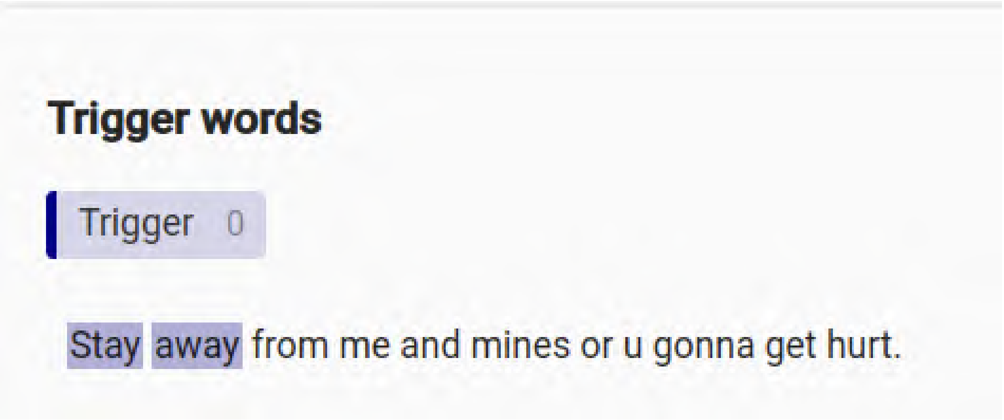
EXALT: multilingual data set for emotion



- Explainability for cross-lingual emotion in tweets
- Trained on English emotion data
- Evaluated on a wide range of target languages (ao Dutch, Russian, Spanish, French, Japanese, …)

Joy 😁
Love 🥰
Fear 😨
Sadness 🙁 | 😭
Anger 😡





GHENT UNIVERSITY

NLP FOR ANCIENT LANGUAGES

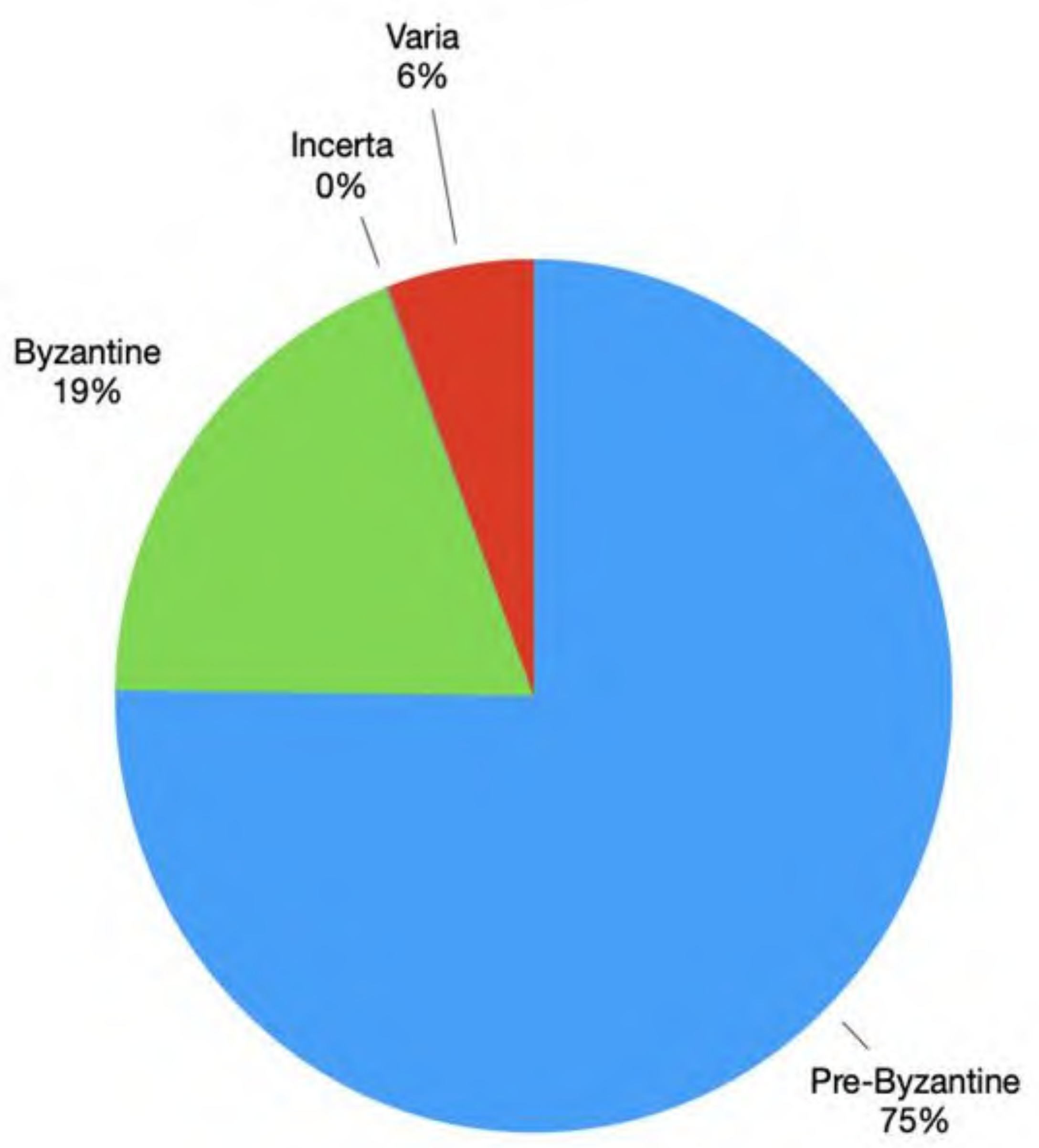
GHENT UNIVERSITY

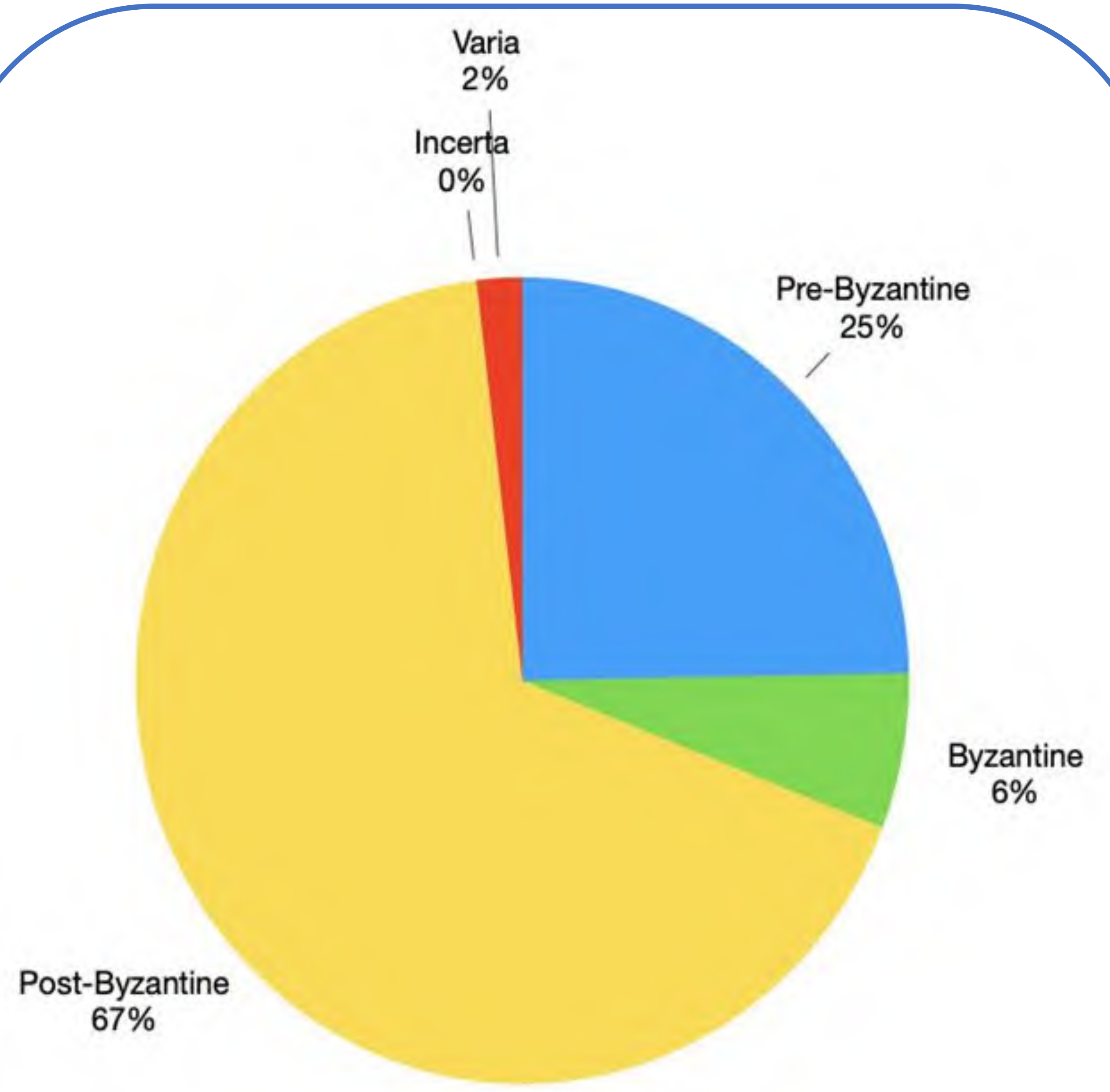

Colin
Swaelens
Pranaydeep
Singh
DBBE

- Interdisciplinary project: “Interconnected texts”: a graph-based computational approach to metrical paratexts in Greek manuscripts (NLP (LT3), Greek literature, Greek linguistics, computer science)
    - NLP: Measuring Orthographic and Semantic Similarity between Byzantine Greek Epigrams

GHENT UNIVERSITY

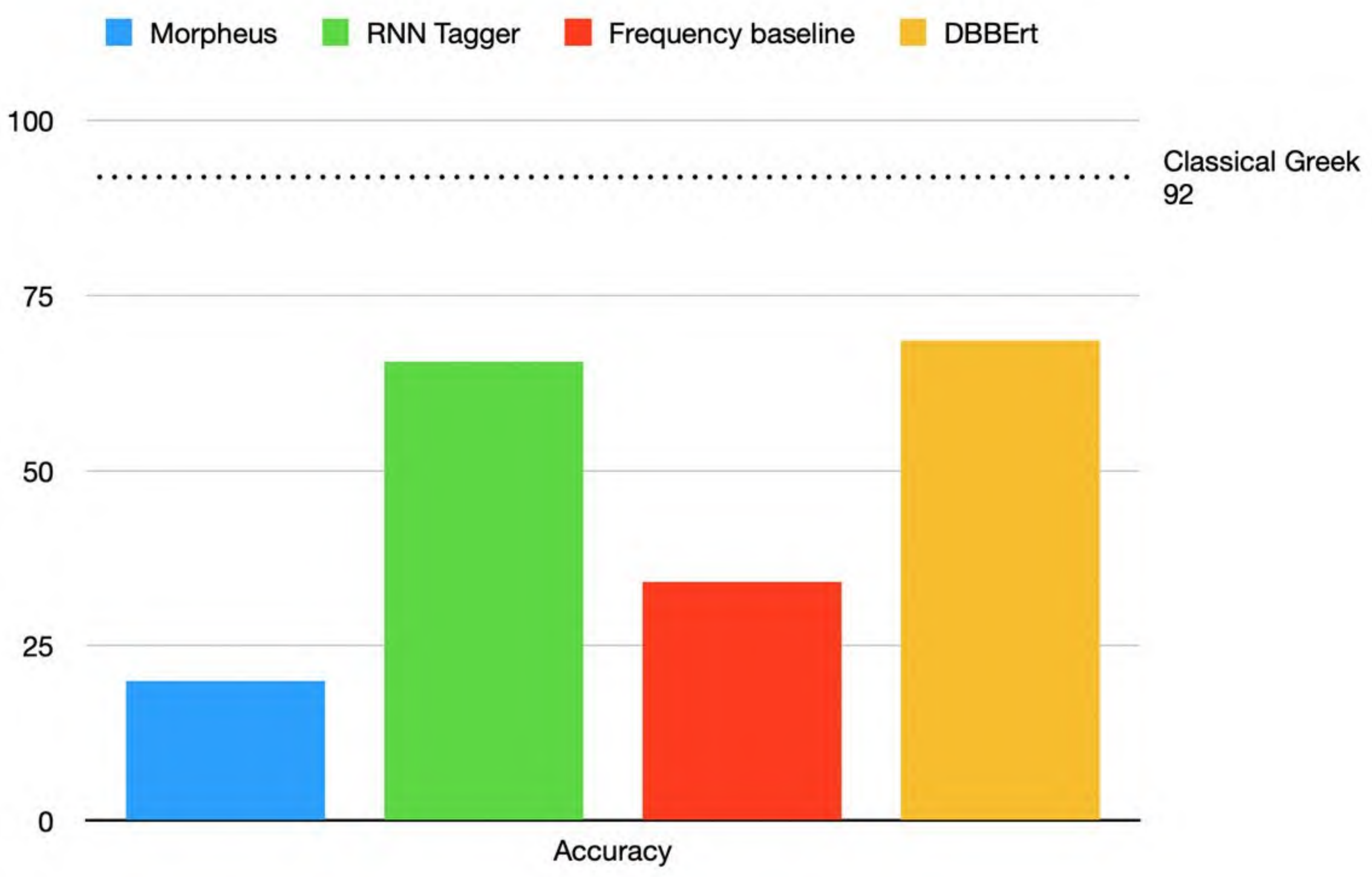
DATA USED TO TRAIN DBBERT





GHENT
UNIVERSITY

Fine-tuning part-of-speech + fine-grained morphology



- Morpheus: rule-based system
- RNN Tagger: best- performing for AG
- Freq. baseline: most occurring label / token
- DBBErt: fine-tuned embedding of our DBBErt

GHENT UNIVERSITY

Gustav Ryberg Smidt

Katrien De Graef

Els Lefever

CUNE-IIIF-ORM

GHENT UNIVERSITY

Towards an Internationally Interoperable
Corpus of Cuneiform Tablets

- IIIF - an image and text API
- OCR - automatically reading cuneiform texts
- **NLP - annotate and analyze Akkadian texts (Ghent University) >** Fully annotated Old Babylonian (c. 2000-1600 B.C.E.) Akkadian letters

GHENT UNIVERSITY

AKKADIAN

- East Semitic language
- Written with the cuneiform script
- In use for more than 2500 years
- Dominated modern-day Iraq

GHENT UNIVERSITY

NLP FOR CUNEIFORM AKKADIAN

ML experiments for Part-of-Speech tagging and morphological annotation:

- Embedding models:
    - Multilingual BERT
    - Semitic PLM: Arabic, Hebrew
    - Japanese

Avg. accuracy results (5-fold on 10K tokens)

PoS (transliterated)

Arabic: 94,1 %

Japanese: 93,4 %

mBERT: 90,3%

PoS + morphological tags

Multilingual: 71,0 %

Arabic: 76,2 %

GHENT UNIVERSITY

NLP FOR CUNEIFORM AKKADIAN

Problems

- Low-resourced language

- Few machine readable texts

- Inconsistent formatting and missing annotation standard

Solutions

> Support with larger Semitic languages (Arabic and Hebrew)

> Specialists gathering data

> Develop UD standards

>> Further investigate impact of:

- different language models / combinations of languages
- adding similar data to train a first Akkadian language model

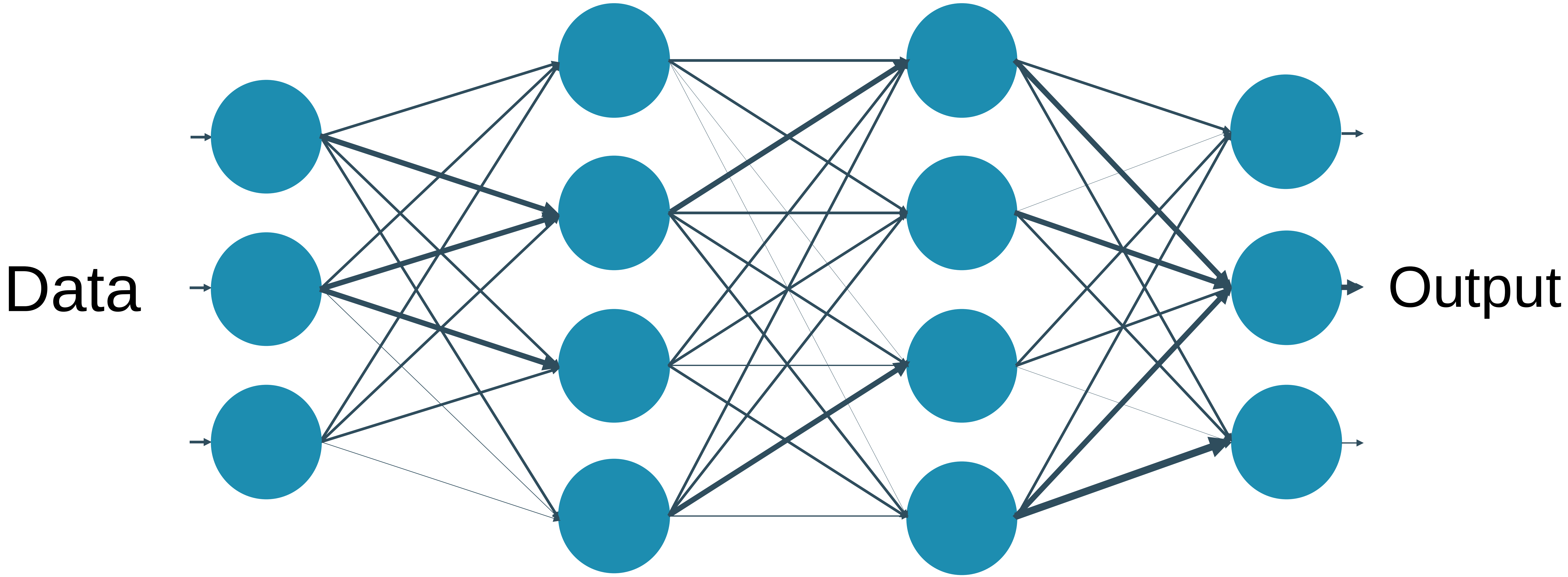
GHENT UNIVERSITY

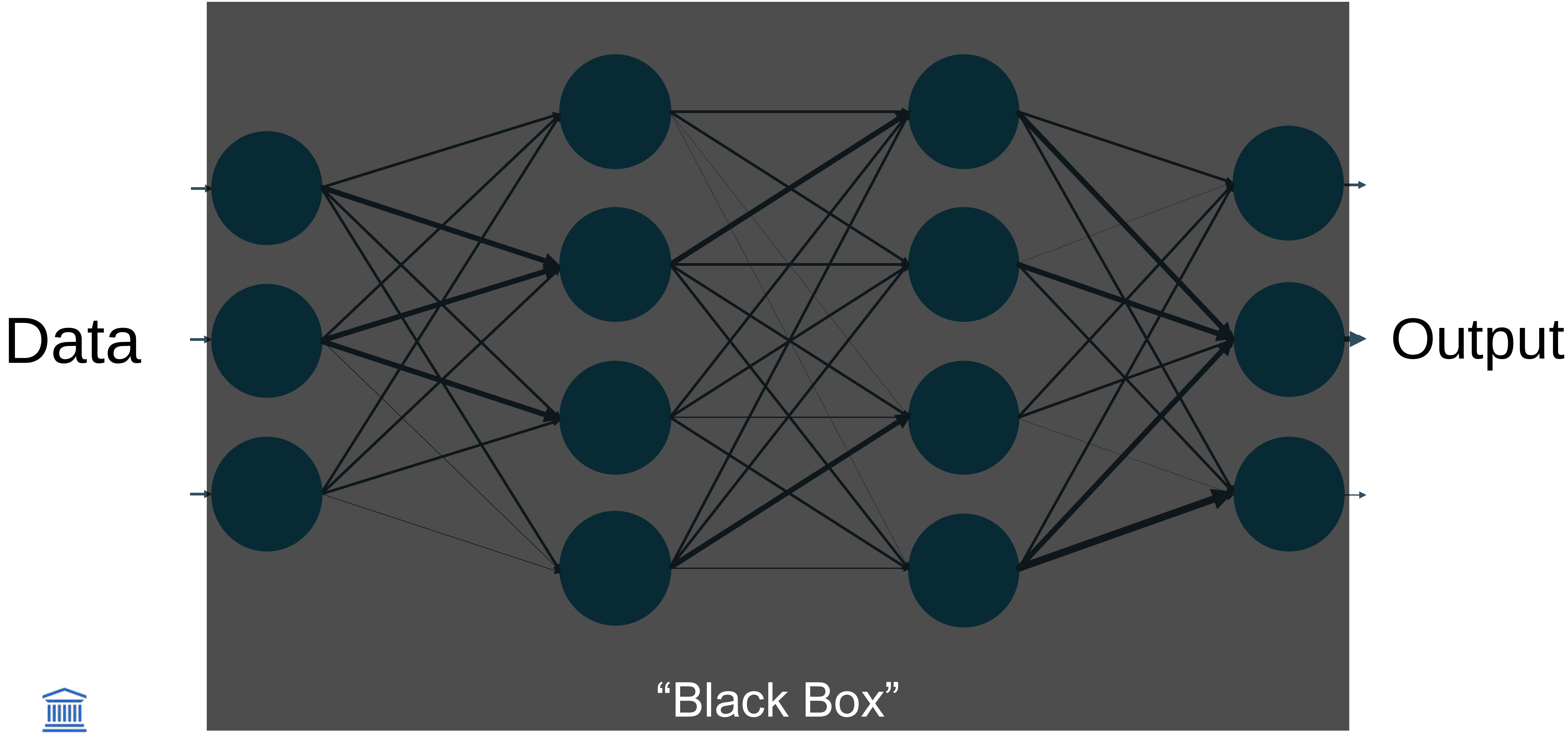
FACULTY OF ARTS
AND PHILOSOPHY

Interpretable NLP systems

GHENT
UNIVERSITY

language and
translation
technology
team


Data
Output


Data
Output
“Black Box”

Aaron Maladry

Els Lefever

Cynthia Van Hee

Véronique Hoste

IRONY DETECTION

GHENT UNIVERSITY

Irony detection

- Manual annotations by trained linguists
- Task: which tweets are ironic and how is the irony realised?

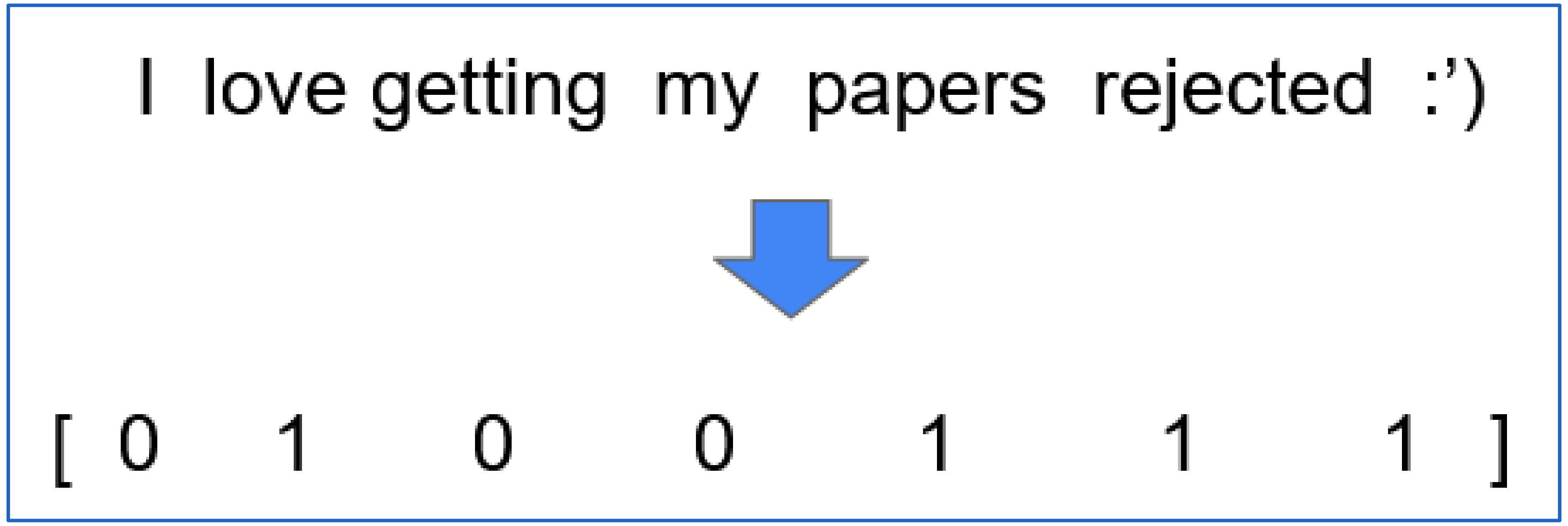
literal sentiment: positive (“can’t wait”)

intended sentiment: negative (“go to the dentist”)

GHENT UNIVERSITY

Irony Detection: trigger words

- trigger word annotation
  - By humans
  - By systems
- advantages:
  - align with system interpretability



GHENT UNIVERSITY

Irony detection: explanations by humans and machines

What do trigger words mean? Why these words? => open to interpretation

- Generate & evaluate explanations
- Compare human and generated explanations

Ironic tweet: *Looooovvveeeeeee when my phone gets wiped*

Explanation: *When your phone gets wiped (which indicates someone did not do it on purpose), you lose all data on your device. This includes a lot of personal information and pictures that people might want to save as keepsakes. As people would not like (accidentally) losing their personal data, the positive evaluation in this tweet is ironic.*

Background knowledge:

- *When a phone gets wiped, all personal data and information is lost.*
- *People do not like losing access to their personal data on their phone*

Irony detection: explanations by humans and machines

Evaluate? Explanation ranking by other group of humans

> works very well for English

> GPT models ranked higher than humans

> other fine-tuned generative explanations are indistinguishable from human explanations

> Next: Dutch explanations !?

Conclusion

Lot of ongoing research and remaining challenges to investigate more fair, robust and interpretable NLP systems:

- carefully curated data sets covering different languages, minority groups, domains, text genres and language variants (historical, dialects, …)
- cross-disciplinary research

GHENT UNIVERSITY

Els.lefever@ugent.be

https://lt3.ugent.be/

GHENT UNIVERSITY