Job matching and fairness

Experiences from a partnership with VDAB

Pieter Delobelle

Workshop on Data-Driven and Algorithmic Tools in Public Employment Services

Independent analysis

Log. regression

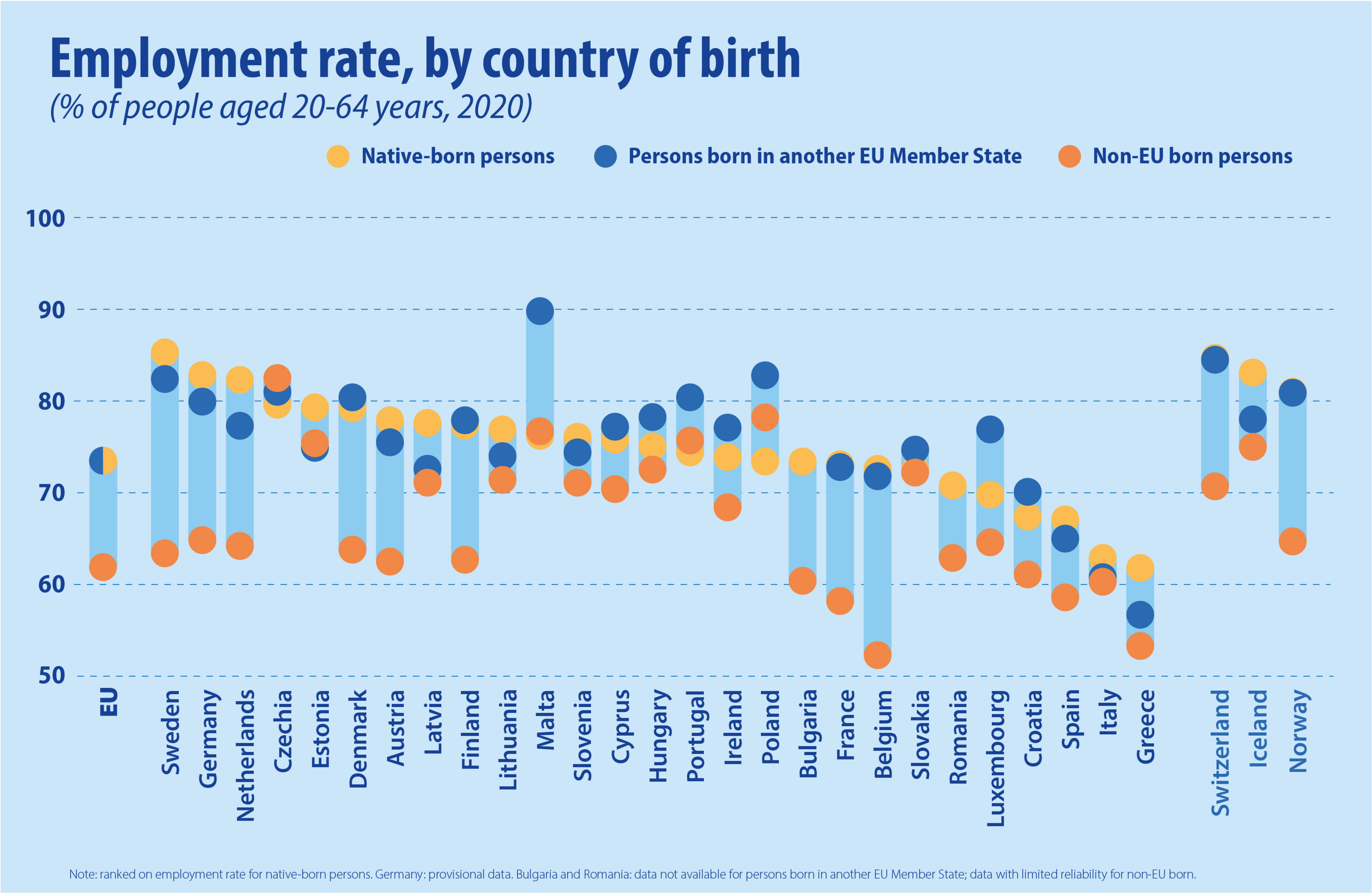
Development

Predictive factors

Proof of concept

NLP (BERT)

Belgium has the lowest employment rate for non-EU persons



ec.europa.eu/eurostat

Partnership with VDAB

- ‘Proof of concept’ in Flemish AI programme
- Research collaboration with Flemish employment service
  - Very unrestricted and research-oriented
  - Access to data to test hypotheses etc…
- Already AI-based systems in use
  - Linking job seekers and vacancies together: **Jobnet**
  - Collaboration to develop new software
  - See Desiere and Struyvens (2019) or an analysis of current systems

DTAI

Desiere, S., & Struyven, L. (2020). Using Artificial Intelligence to classify Jobseekers: The Accuracy-Equity Trade-off.

VDAB

Partnership with VDAB

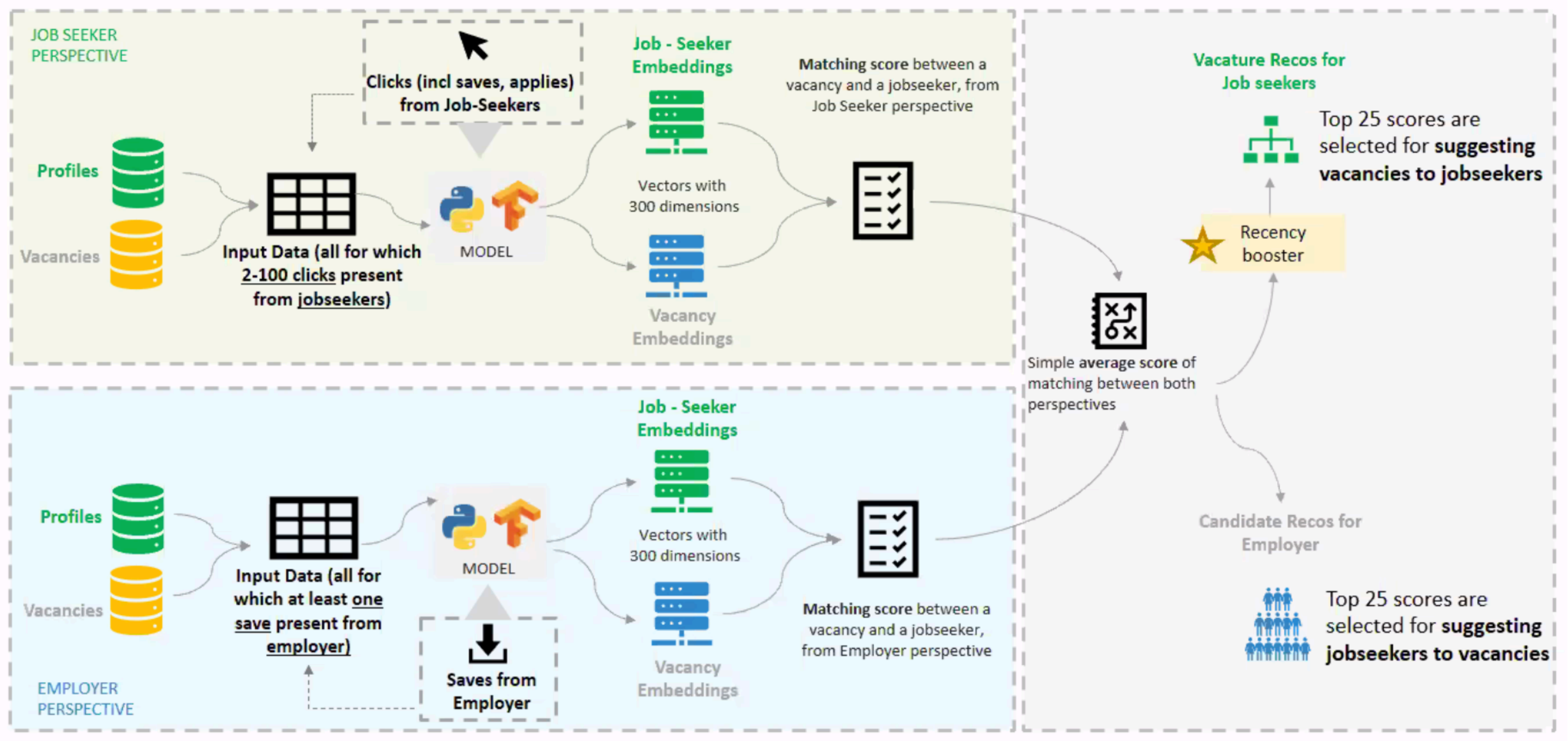
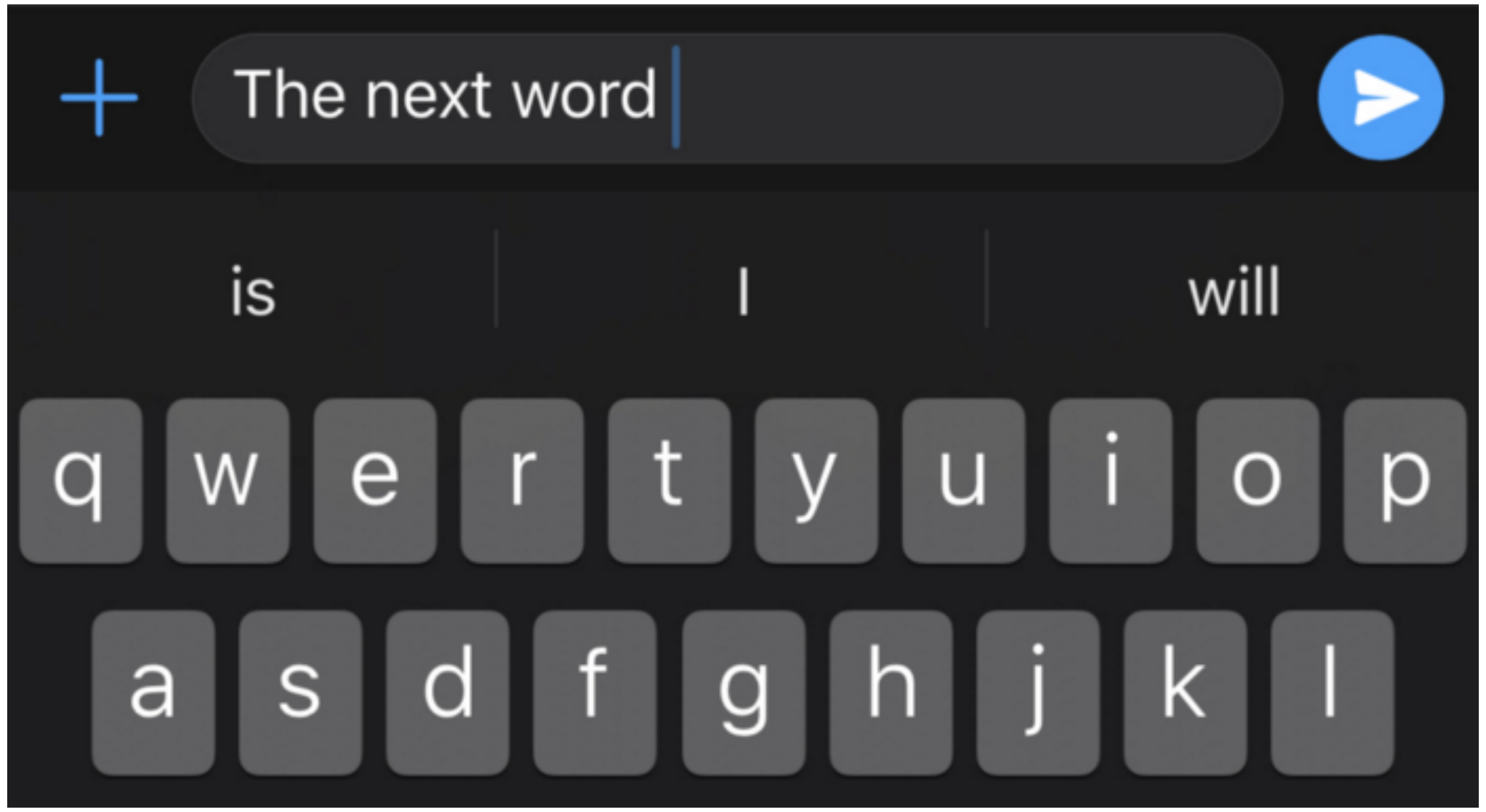
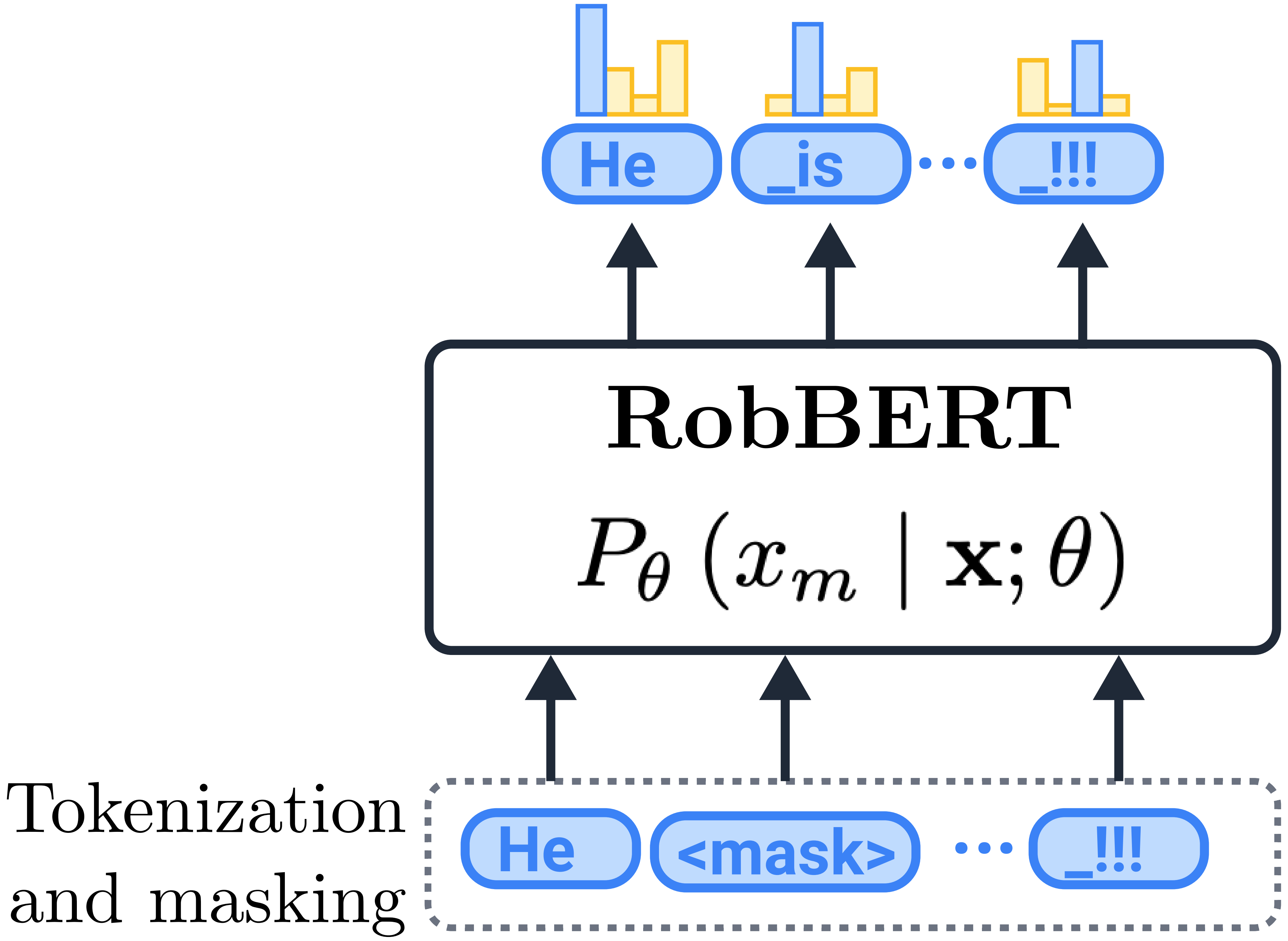


Figure by Vincent Buekenhout (2020).

Our research

Language modeling





Causal language modeling

GPT-2, GPT-3

Masked language modeling (MLM)

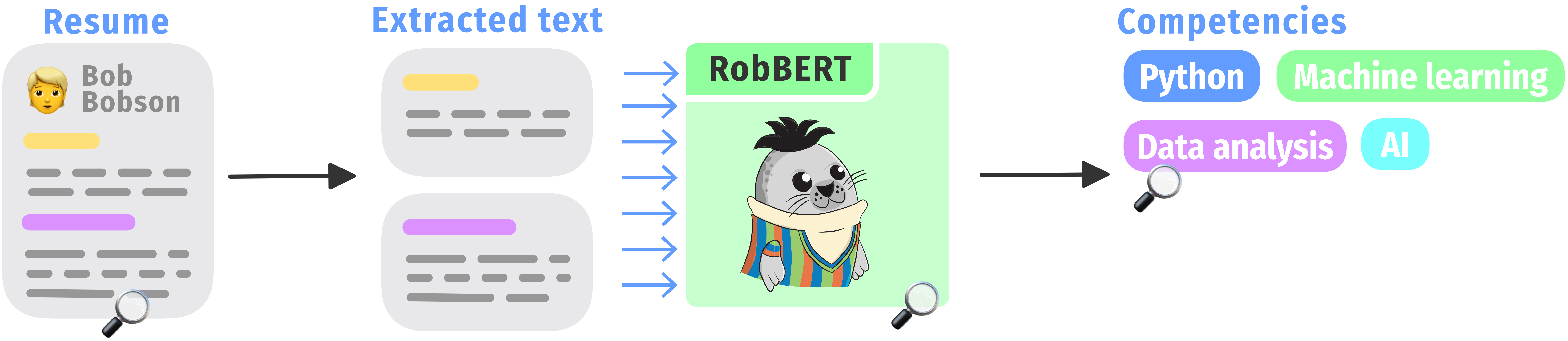
BERT, RoBERTa, RobBERT, ...

DTAI

Extracting features from text

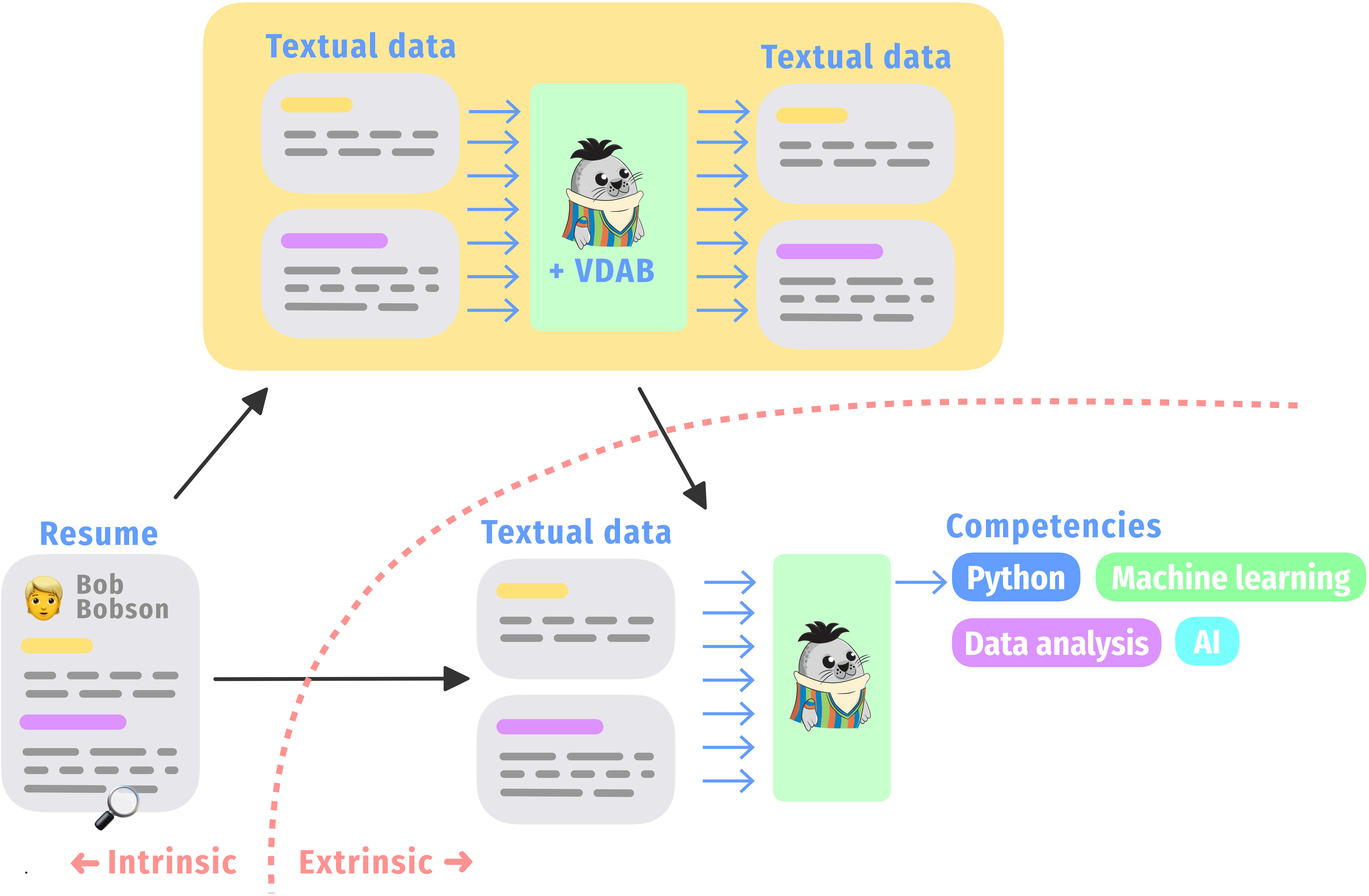
Goal

Using language models to extract features from unstructured data (e.g. text from resumes, ..). These **extracted features will then be used for a fairness evaluation**, for instance when linking resumes to skills.



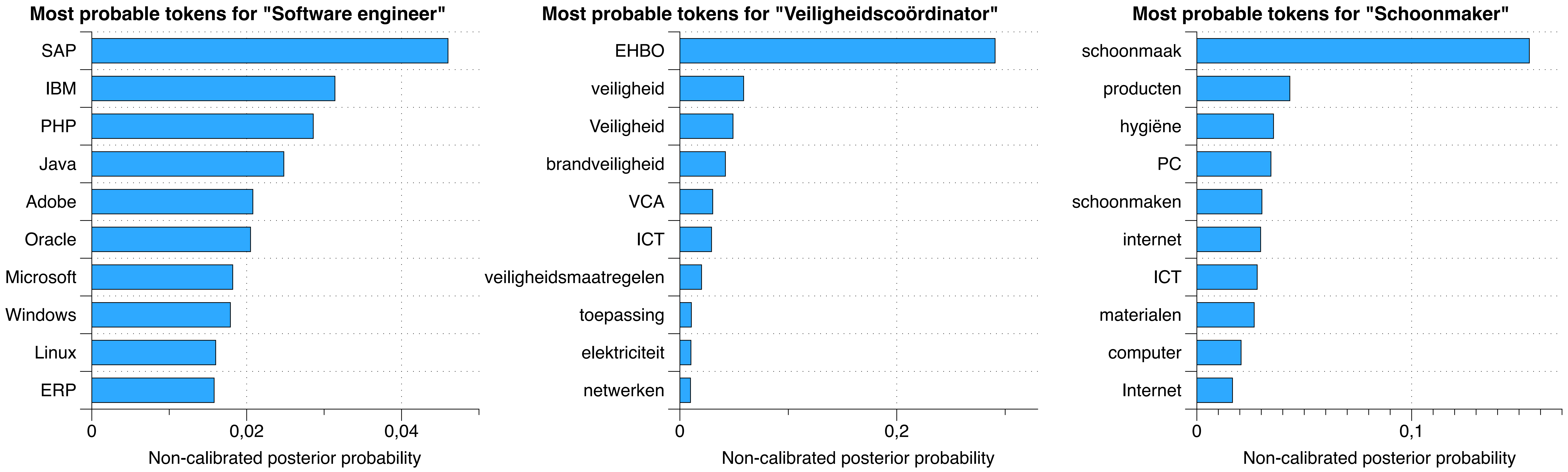
DTAI

Using domain adapted language models



DTAI

Predicting skills based on profession titles



DTAI

Our research

What is fairness in PES?

Gender stereotyping — Do LM fairness metrics make sense?

Discovery of correlations (DisCo)

$$p_{tgt} = P\left(X_m = t \mid \mathbf{x}; \theta\right)$$

Target t is usually a pronoun (‘He’, ‘She’)

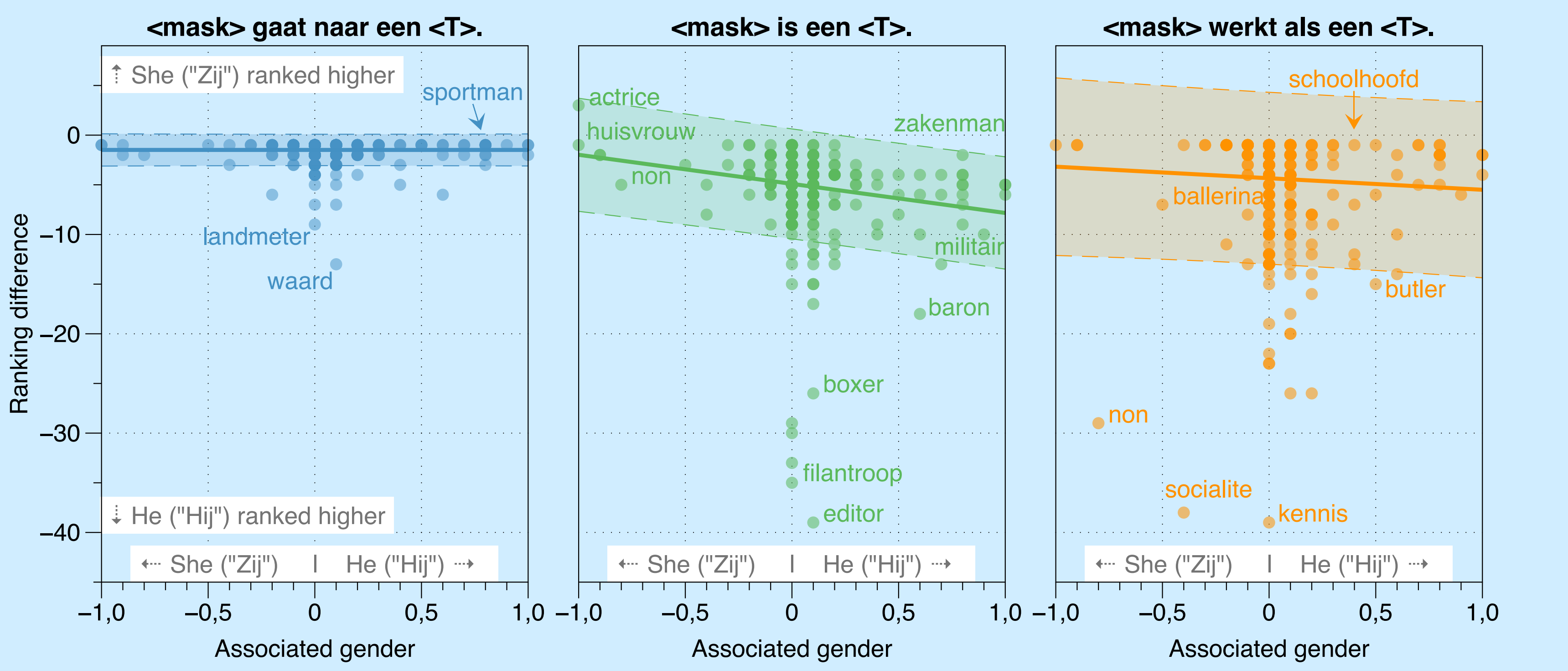
Log probability bias score

$$p_{prior} = P\left(X_m = t \mid \mathbf{x} \backslash \{x_p\}; \theta\right)$$

$$\log \frac{p_{tgt}}{p_{prior}}$$

Difference between associations of two targets

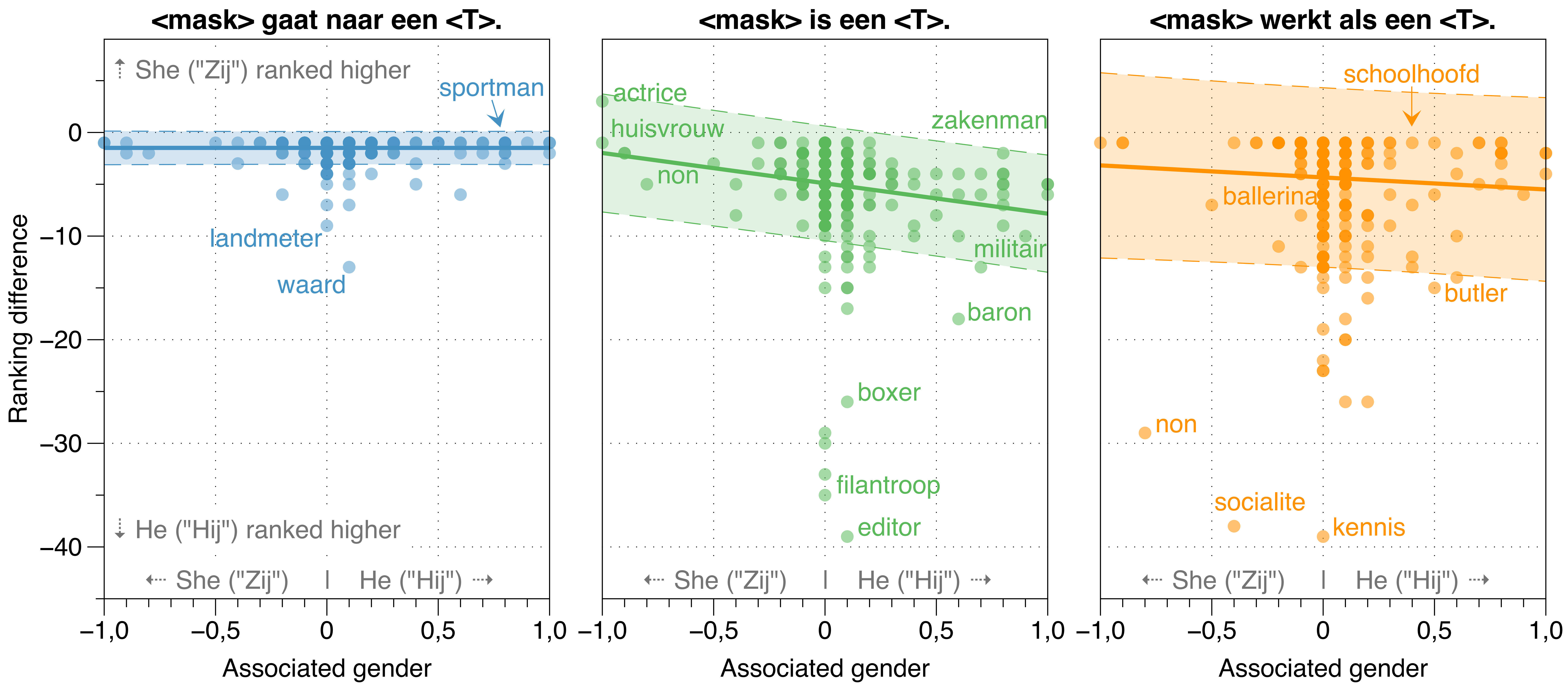
Gender stereotyping in RobBERT



Webster et al. 2020. ‘Measuring and Reducing Gendered Correlations in Pre-Trained Models’.
Delobelle et al. 2020. ‘RobBERT: A Dutch RoBERTa-Based Language Model’.

Gender stereotyping

Third person pronouns commonly used, **does this make sense for résumés?**

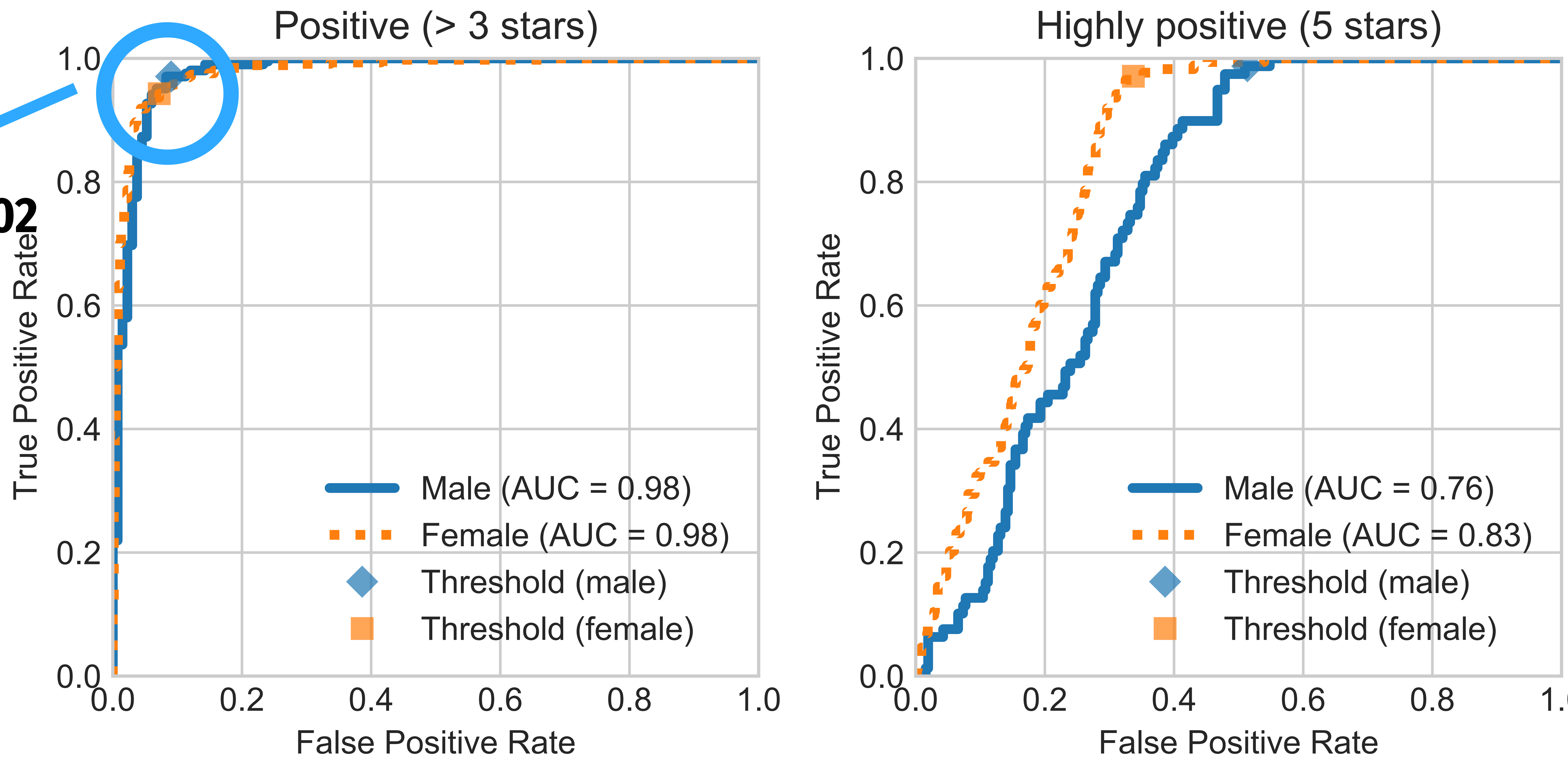


DTAI

Delobelle, Pieter, Thomas Winters, and Bettina Berendt. 2020. 'RobBERT: A Dutch RoBERTa-Based Language Model'.

This affects predictions

- Evaluate sensitive attributes on Dutch book reviews
- “RobBERT thinks women are more predictable than men”



Demographic parity ratio: **0.702**

Equal opportunity: **0.028**

Delobelle, Pieter, Thomas Winters, and Bettina Berendt. 2020. ‘RobBERT: A Dutch RoBERTa-Based Language Model’.

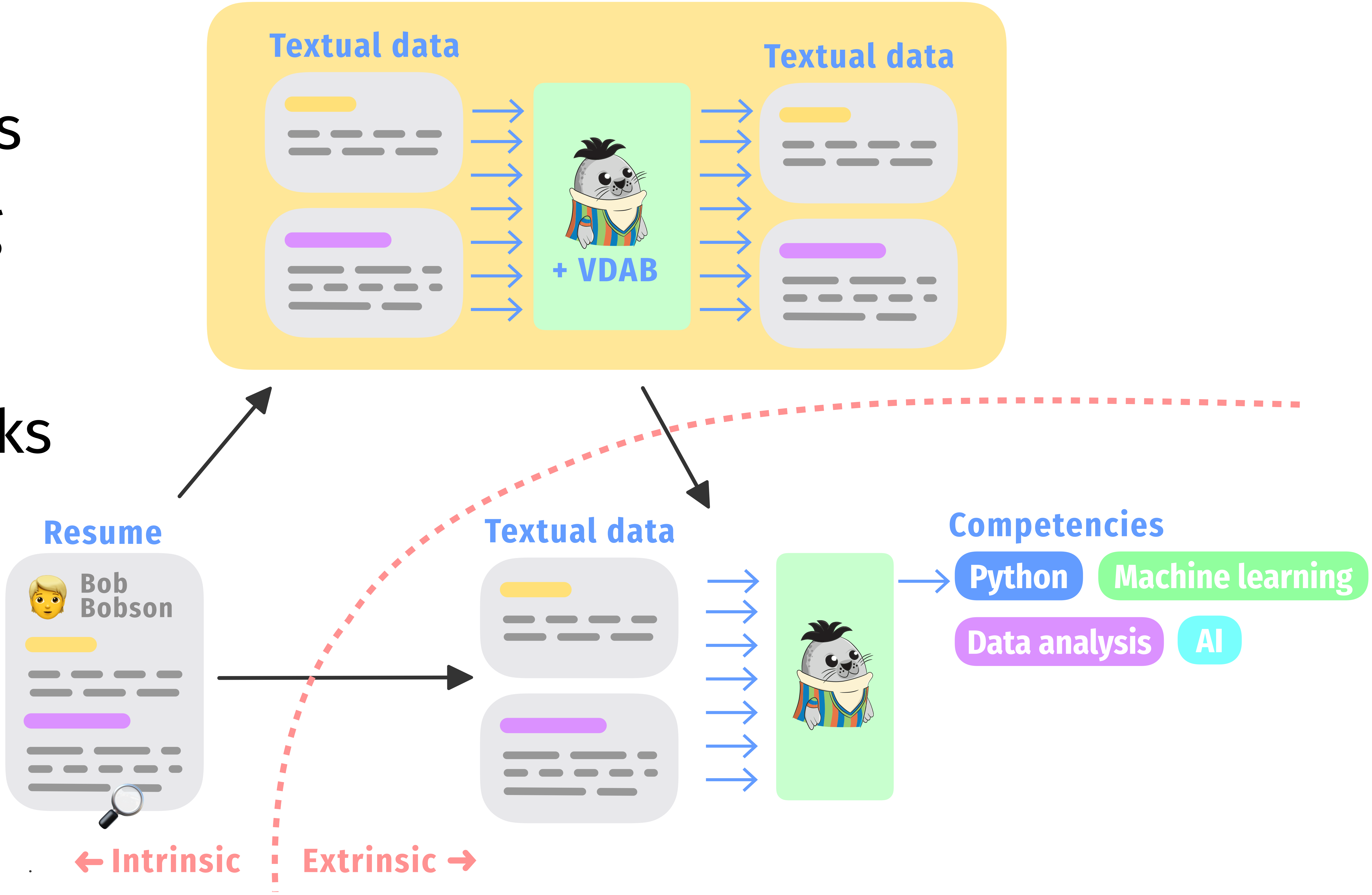
Sources of bias —pre-training and finetuning

Intrinsic measures

- Bias in language models
- eg. gender stereotyping

Extrinsic measures

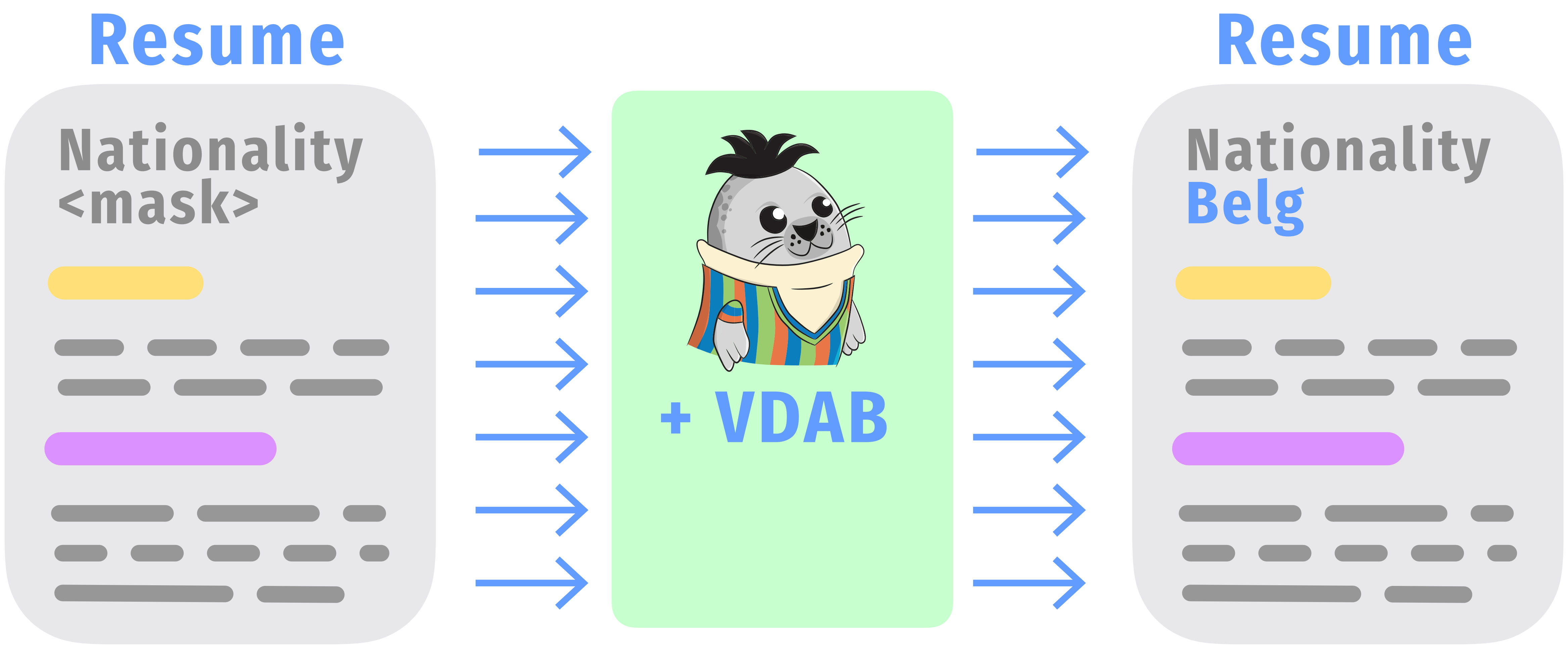
- Bias in downstream tasks
- Resource allocations
- Real harms



Webster et al. 2020. 'Measuring and Reducing Gendered Correlations in Pre-Trained Models'.

Using context from CVs

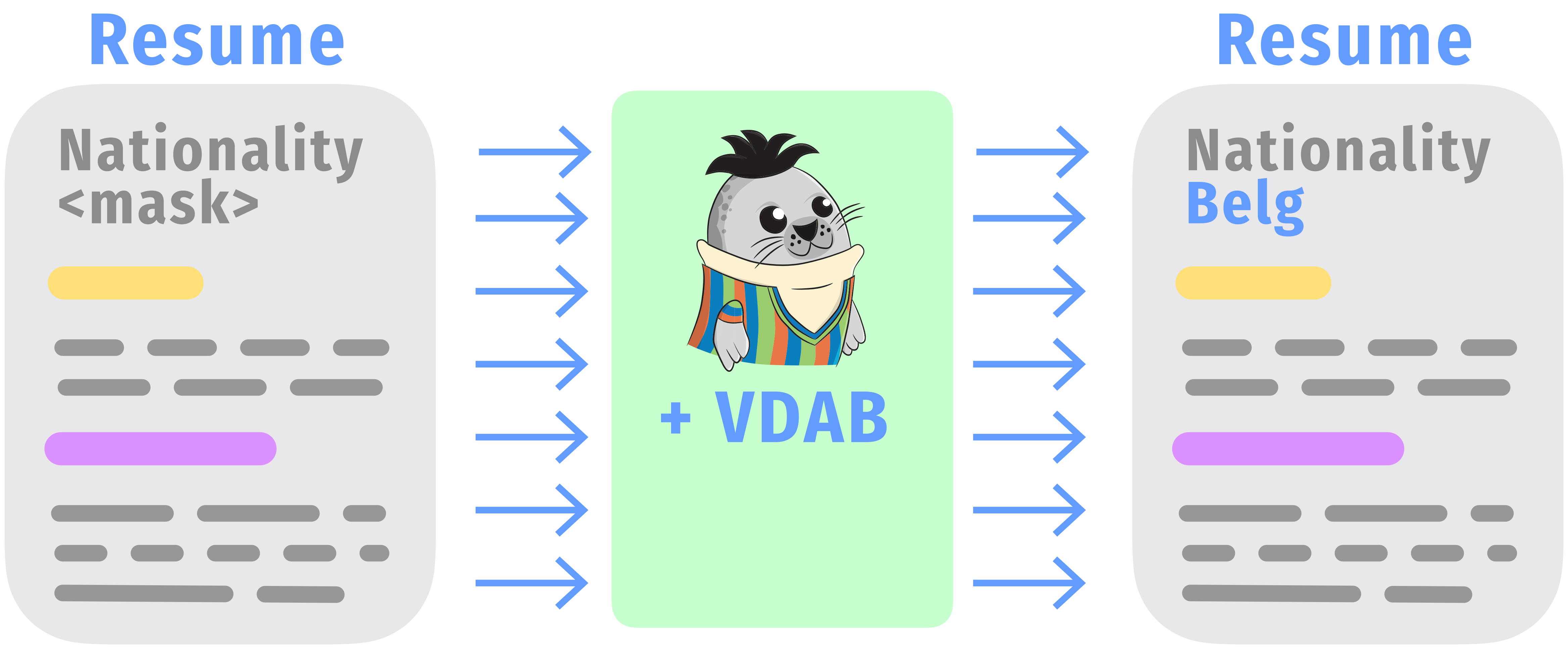
- Leverage MLM task to predict protected attributes
- This is a **contextualized prediction** given the resume

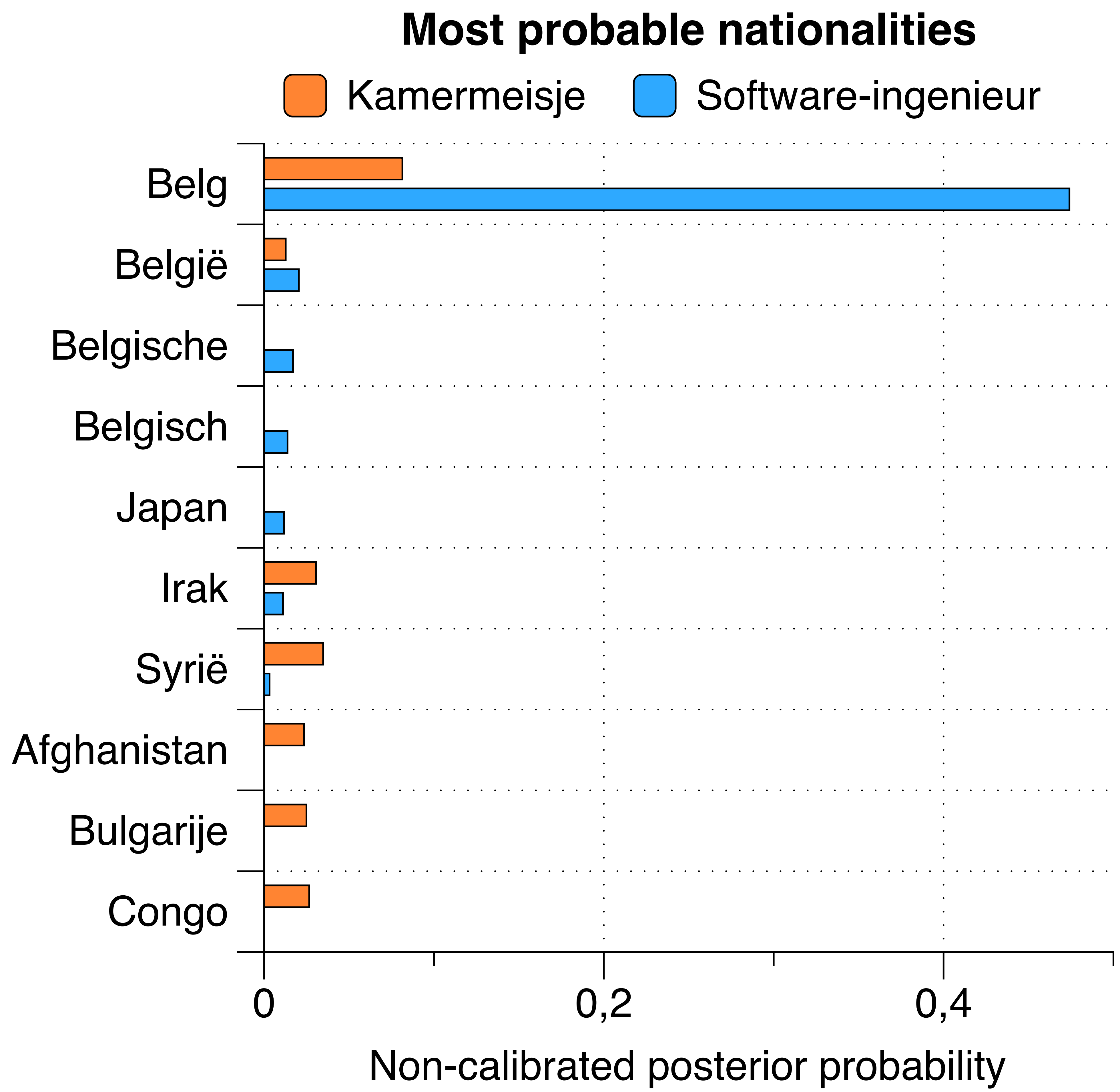


DTAI

Using context from CVs

- Leverage MLM task to predict protected attributes
- This is a **contextualized prediction** given the resume
- "Cleaning ladies aren't Belgians"

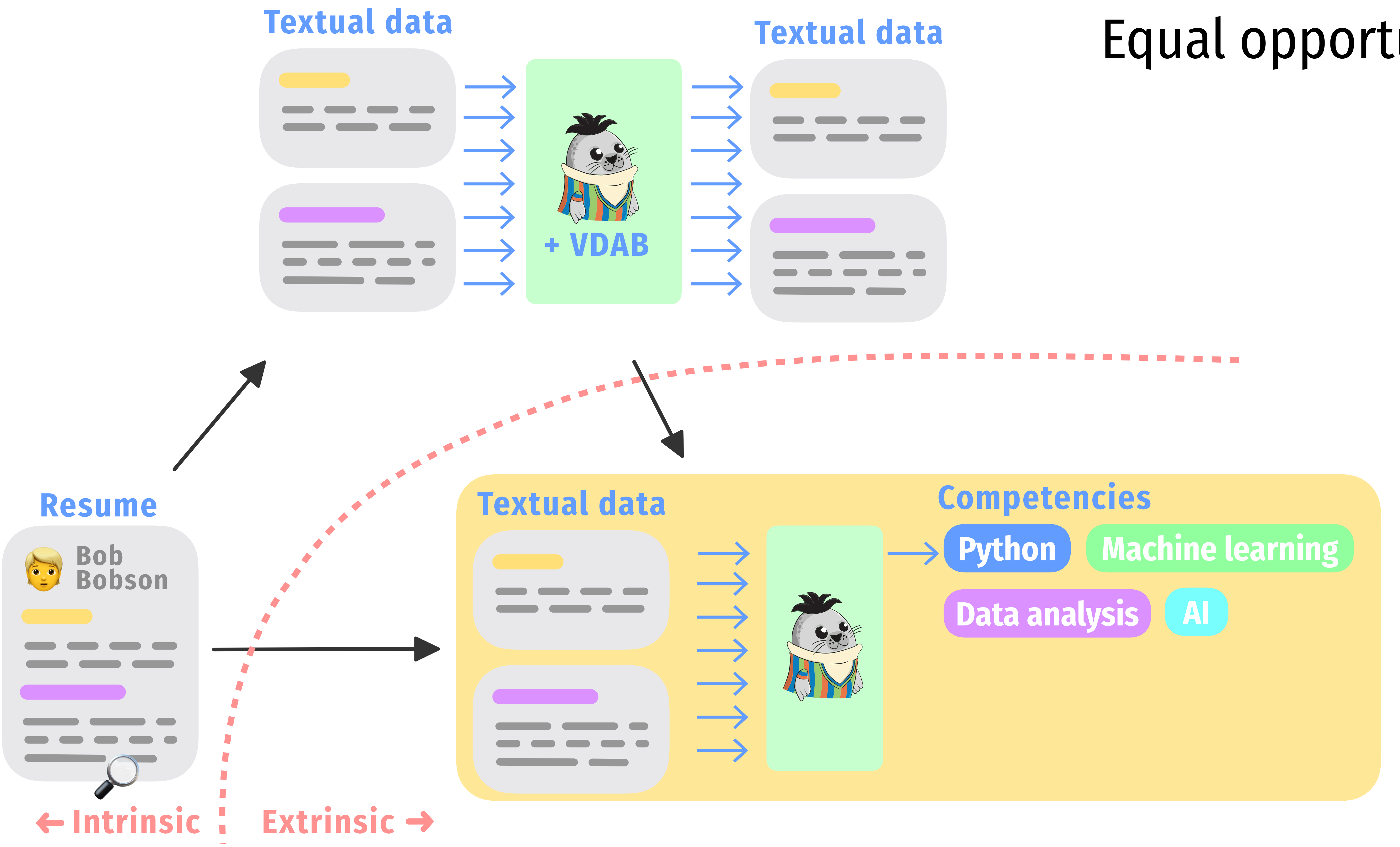




DTAI

What’s next?

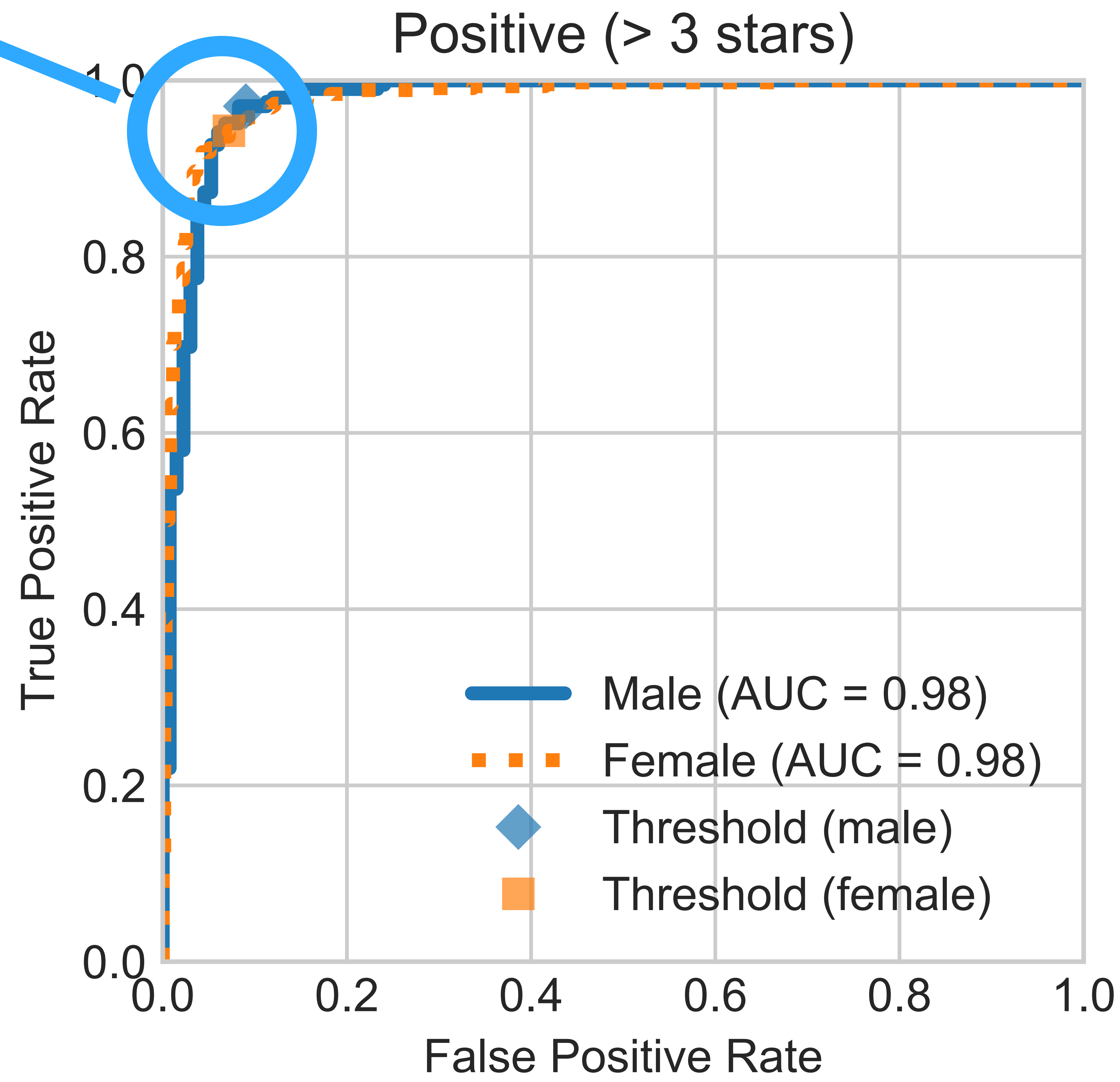
Stereotyping in LMs **might** create harms in downstream tasks



Dutch Book Reviews dataset

Demographic parity ratio: **0.702**

Equal opportunity: **0.028**



DTAI

Delobelle et al. 2020. 'RobBERT: a Dutch RoBERTa-based Language Model'

Thank you!

pieter.delobelle@kuleuven.be

DTAI
Declarative Languages & Artificial Intelligence

KU LEUVEN