

Footprint of Societal Biases in Natural Language Processing



Navid Rekab-Saz

navid.rekabsaz@jku.at

Agenda

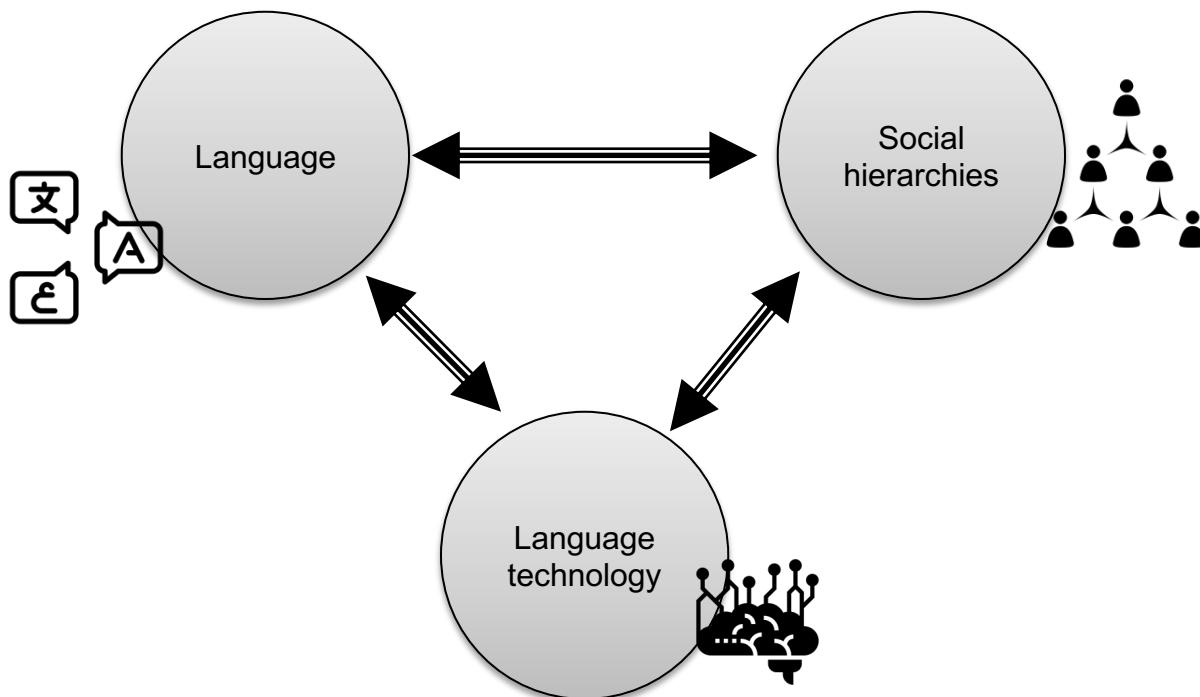
- Societal biases in NLP ... what? why?
- Bias in word embedding
- Measuring bias in Information Retrieval

Agenda

- **Societal biases in NLP ... what? why?**
- Bias in word embedding
- Measuring bias in Information Retrieval

Language and Society

- Language ...
 - takes on and defines social meaning
 - forms and maintains social hierarchies by ...
 - labeling social groups
 - transmitting the beliefs about social groups



(Societal) Bias

- Biases and stereotypes do not *per se* carry negative meanings!
- What we mean from “bias” in this lecture is ...

“Inclination or prejudice for or against one person or group, especially in a way considered to be unfair.”

Oxford dictionary

“demographic disparities in algorithmic systems that are objectionable for societal reasons.”

Fairness and Machine Learning

Solon Barocas, Moritz Hardt, Arvind Narayanan, 2019, fairmlbook.org

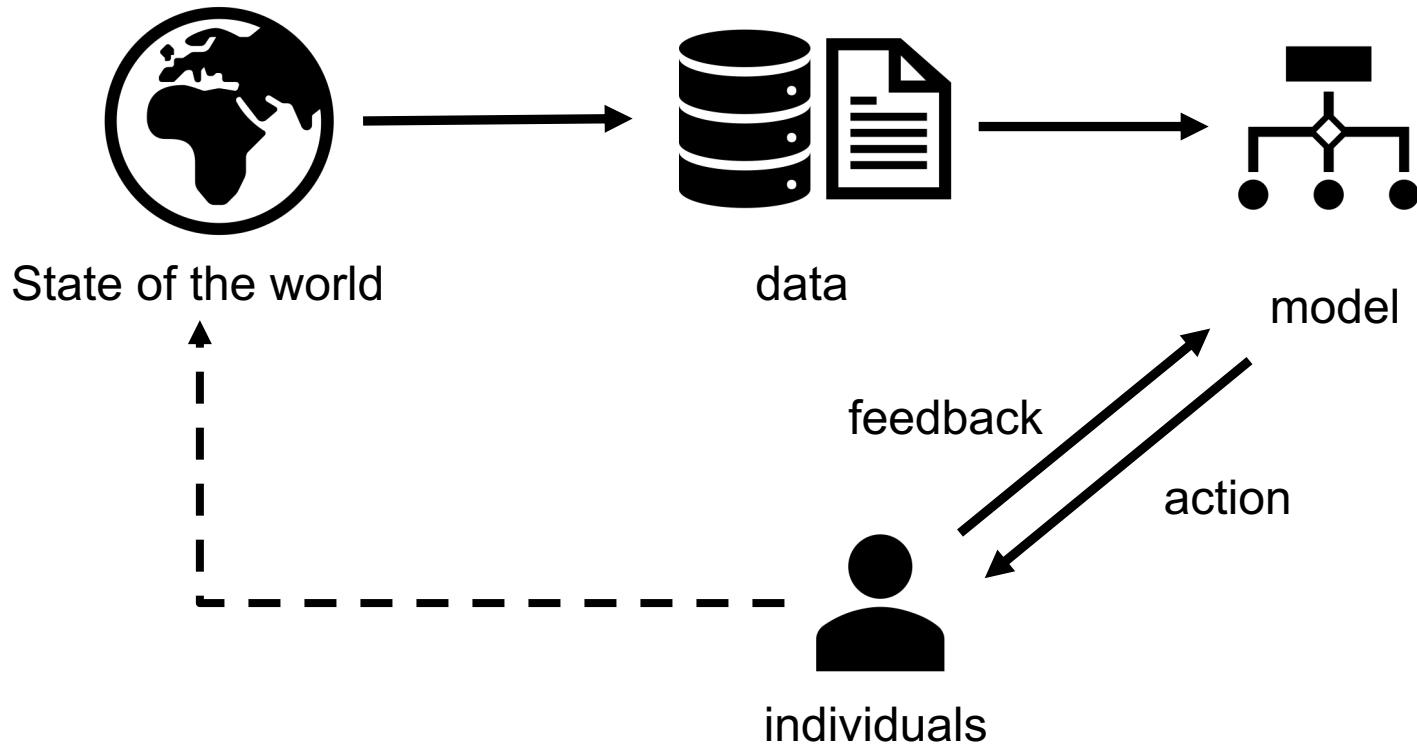


"I think your test grading is biased in favor of students who answer the test questions correctly."

In which ways can bias be *harmful*?

- Allocational harms
 - A system allocates resources (e.g., credit) or opportunities (e.g., jobs) unfairly to different social groups
- Representational harms
 - a system (e.g., a search engine or a recommender system) represents some social groups in a less favorable light than others.
 - For instance by stereotyping that propagates negative generalizations about particular social groups

Machine Learning Cycle



Bias in image processing

Google says sorry for racist auto-tag in photo app

<https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>

FaceApp's creator apologizes for the app's skin-lightening 'hot' filter

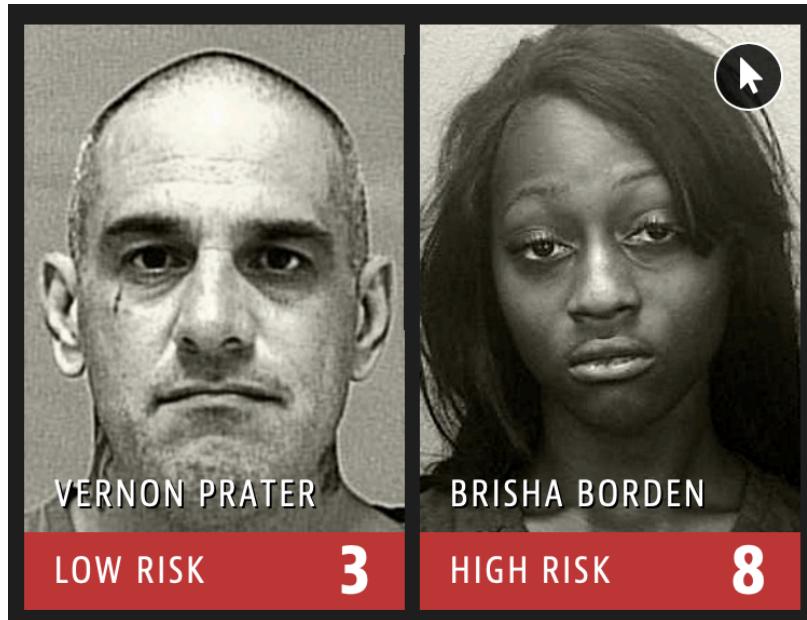
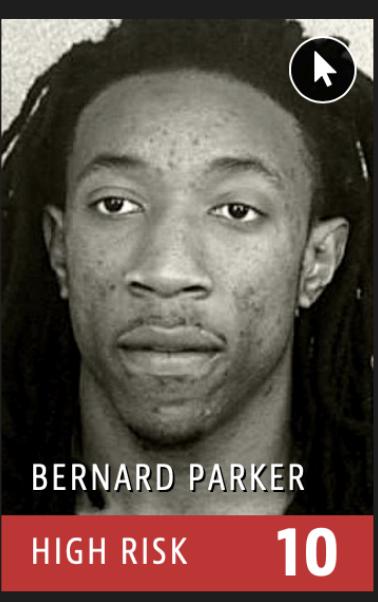
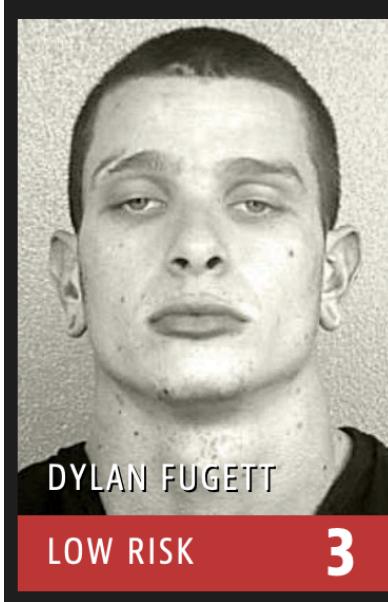
<https://www.theverge.com/2017/4/25/15419522/faceapp-hot-filter-racist-apology>

Beauty.AI's 'robot beauty contest' is back – and this time it promises not to be racist

<https://www.wired.co.uk/article/robot-beauty-contest-beauty-ai>

Bias in crime discovery

- Predicted risk of reoffending



<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Bias in Image Retrieval

Search: nurse



All

Images

News

Videos

Maps

More

Settings

Tools

Collections

SafeSearch



Foreign-Educated Nurse
nurse.org



documenting properly on patient charts
nurse.com



One Nurse at a Time Helps ...
online.nursing.georgetown.edu



Mountain Valleys Health Centers
mtnvalleyhc.org



Nurse Staffing
nursingworld.org



Magic Back to Nursing in 2019
nurse.org



Changes to Nurse Licensure Compact ...
taledmed.com



Nursing Jobs
nursingworld.org



Practical Nurse Program - A...
abccott.edu



Auto Insurance for Nurses from ...
mycalcas.com



An Irish nurse: 'I began to hate my job ...
thejournal.ie



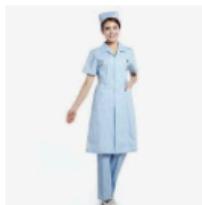
Nurse interview questions and answers ...
snagajob.com



an Emergency Room Nurse
americanr.com



Men in Nursing Archives - Minority Nurse
minoritynurse.com



Hospital Nurse Uniform at Rs...
indiamart.com



Johnson Looking for Nurse Innovators ...
elitemma.com



IoT vendors. When a paediatric nurse ...
theranister.co.uk



know if Nursing is Your Dream Job ...
advanced-care.us

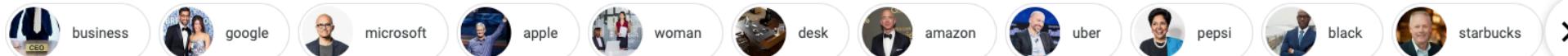
Bias in Image Retrieval

Search: CEO



All Images News Maps Videos More Settings Tools

Collections SafeSearch ▾



Chief executive officer - Wikipedia
en.wikipedia.org



Odilon Almeida as President and Chief ...
businesswire.com



You are the CEO of Your Life | Personal ...
personalexcellence.co



What do CEOs do? A CEO Job Descriptio...
steverobbins.com



CEO vs. Owner: The Key Differences ...
onlinemasters.ohio.edu



Why You Need To Be The CEO Of Your C...
forbes.com



How to use 'CEO magic' when trying ...
europeanceo.com



LinkedIn CEO Jeff Weiner steps down ...
fortune.com



ABB ernennt neuen CEO | Netzwoche
netzwoche.ch



This South African has been named amon...
businesstech.co.za



CEO of the Year
chiefexecutive.net



Harvard study: What CEOs do all day
cnbc.com



C.E.O. Fired Over a Relationship ...
nytimes.com



ESSENCore
essencore.com



How CEOs can help lead technology ...
mckinsey.com



Becoming a CEO: Unexpected traits top ...
execed.economist.com



Casey's Announces CEO Tr...
businesswire.com



CEO MESSAGE | JCB Global Website
global.jcb

Bias in Machine Translation

The screenshot shows a machine translation interface with the following components:

- PERSIAN - DETECTED**: The source language is detected as Persian.
- PERSIAN**: The input text in Persian.
- ENGLISH**: The target language is English.
- SPANISH**: An additional language option.
- Translations**: The Persian input is translated into English, but the gender pronouns are consistently swapped between male and female forms.
- Feedback**: A red arrow points from the bottom text "same gender-neutral pronoun" to the gendered English translations, highlighting the lack of neutrality.
- Metrics**: A progress bar at the bottom indicates 86/5000, and a red arrow points down to the text "same gender-neutral pronoun".
- Icons**: Standard icons for audio, copy, and more options.

Key observations:

- The Persian input consists of seven sentences, each starting with "او" (Av) followed by a gendered title (Manager, Nurse, Doctor, Beautiful, Cute, Funny, Genius).
- The English output consistently swaps the gender pronouns ("He" for female and "She" for male), demonstrating a clear bias.
- A red arrow points from the text "same gender-neutral pronoun" at the bottom to the gendered English translations, emphasizing the lack of neutrality in the current system.

Text in Persian input:

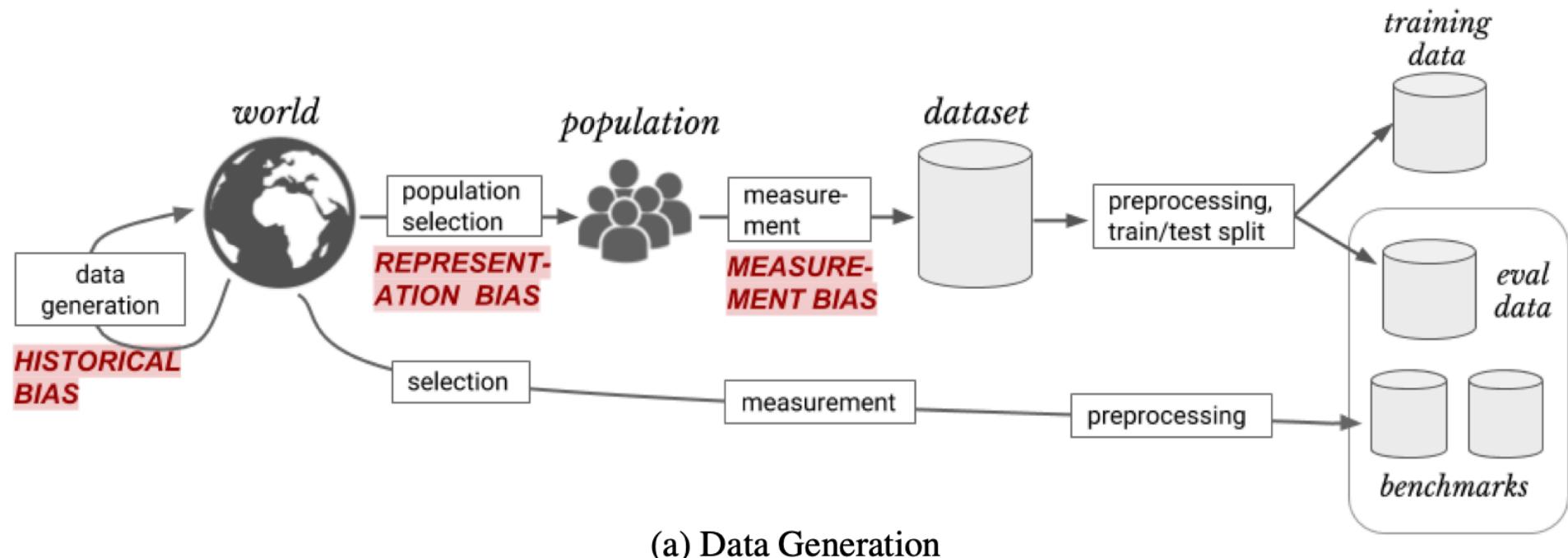
- او مدیر است
- او پرستار است
- او دکتر است
- او زیبا است
- او ناز است
- او بامزه است
- او نابغه است

Text in English output:

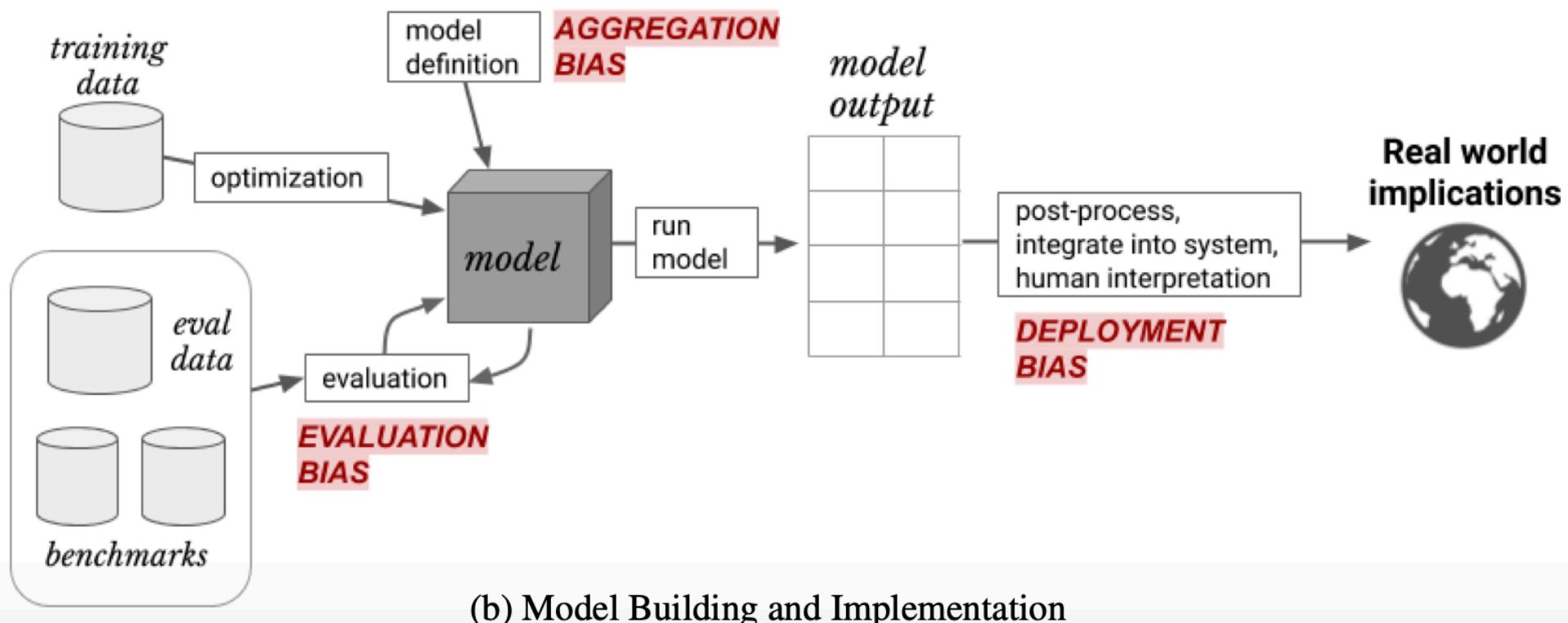
- He is the manager
- She is a nurse
- He is a doctor
- She is beautiful
- She is cute
- He is funny
- He is a genius

Bottom text: same gender-neutral pronoun

Bias in Machine Learning



Bias in Machine Learning



Where is bias originated from?

- World (**historical bias**)
 - Historical and ongoing discrimination
- Data (**representation bias / measurement bias**)
 - Sampling strategy - who is included in the data?
- Models (**aggregation bias**)
 - Using sensitive information (e.g. race) directly or adversely
 - Naive modeling learns more accurate predictions for majority group
 - Algorithm optimization eliminates “noise”, which might constitute the signal for some groups of users
- Evaluations (**evaluation bias**)
 - Definition of Success
 - Who is it good for, and how is that measured? Who decided this? To whom are they accountable?
 - Data annotation and benchmarking
- Human interaction (**deployment bias**)

Bias & Fairness in standard Machine Learning

Attributes

- • age
- workclass
- fnlwgt
- education
- marital-status
- occupation
- relationship
- • race
- • sex
- capital-gain
- capital-loss
- hours-per-week
- native-country



whether a person makes over 50K a year

```
39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Male, 2174, 0, 40, United-States, <=50K
50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 13, United-States, <=50K
38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 40, United-States, <=50K
53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0, 40, United-States, <=50K
28, Private, 338409, Bachelors, 13, Married-civ-spouse, Prof-specialty, Wife, Black, Female, 0, 0, 40, Cuba, <=50K
37, Private, 284582, Masters, 14, Married-civ-spouse, Exec-managerial, Wife, White, Female, 0, 0, 40, United-States, <=50K
49, Private, 160187, 9th, 5, Married-spouse-absent, Other-service, Not-in-family, Black, Female, 0, 0, 16, Jamaica, <=50K
52, Self-emp-not-inc, 209642, HS-grad, 9, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 45, United-States, >50K
31, Private, 45781, Masters, 14, Never-married, Prof-specialty, Not-in-family, White, Female, 14084, 0, 50, United-States, >50K
42, Private, 159449, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 5178, 0, 40, United-States, >50K
37, Private, 280464, Some-college, 10, Married-civ-spouse, Exec-managerial, Husband, Black, Male, 0, 0, 80, United-States, >50K
30, State-gov, 141297, Bachelors, 13, Married-civ-spouse, Prof-specialty, Husband, Asian-Pac-Islander, Male, 0, 0, 40, India, >50K
23, Private, 122272, Bachelors, 13, Never-married, Adm-clerical, Own-child, White, Female, 0, 0, 30, United-States, <=50K
```

Bias & Fairness in NLP

- In language, bias can hide behind the implicit meanings of words and sentences

A sample task – occupation prediction from biographies:

[She] graduated from Lehigh University, with honours in 1998.
[Nancy] has years of experience in weight loss surgery, patient support, education, and diabetes



Nurse

Real problems need interdisciplinary thinking!

- Fairness and bias are **social concepts** and inherently **normative**
- Bias in NLP systems should be grounded in its **social context**

“... without this grounding, researchers and practitioners risk measuring and mitigating only what is convenient to measure and mitigate, rather than what is most normatively concerning.”

Blodgett et al. [2020]

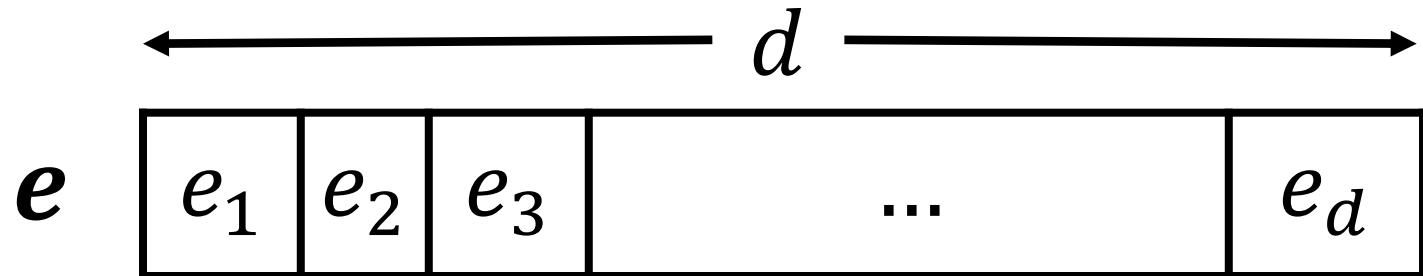
- Addressing bias in NLP requires going beyond CS and getting engaged with disciplines such as sociolinguistics, linguistic anthropology, sociology, law, psychology, etc..

Agenda

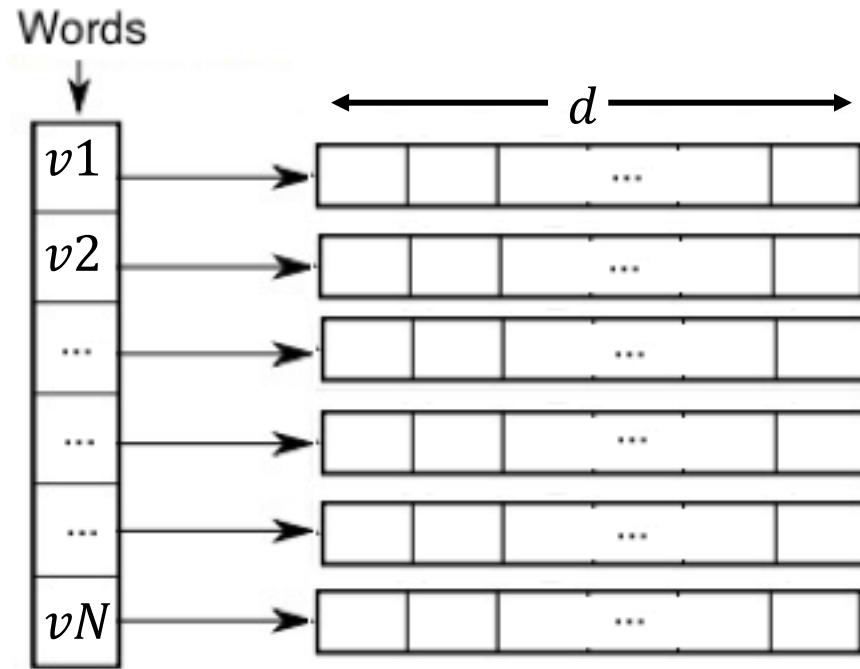
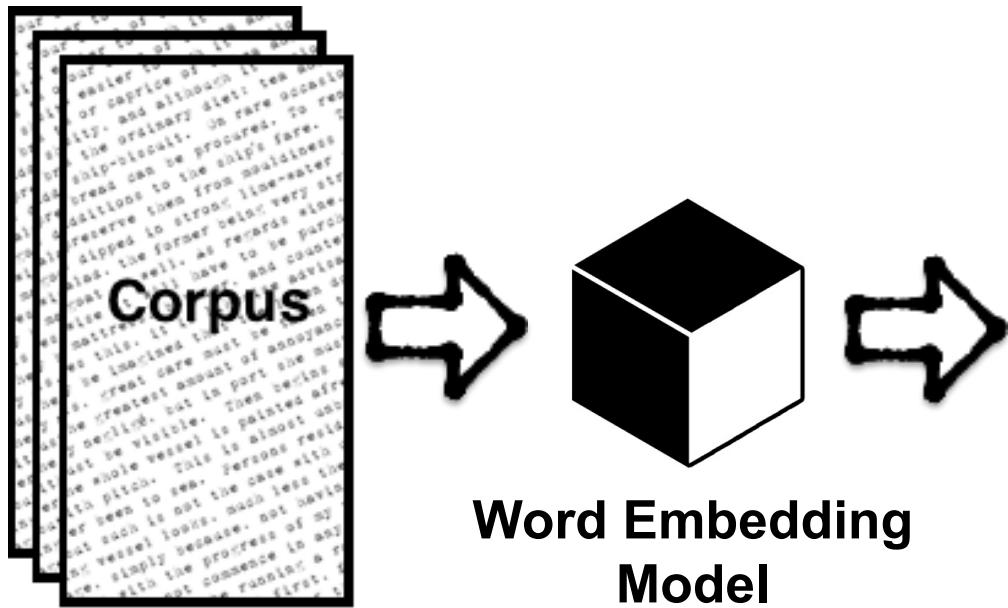
- Societal biases in NLP ... what? why?
- **Bias in word embedding**
- Measuring bias in Information Retrieval

Word Embedding

- A word is represented with a vector of d dimensions
- The vector represents the meaning or semantics of the word



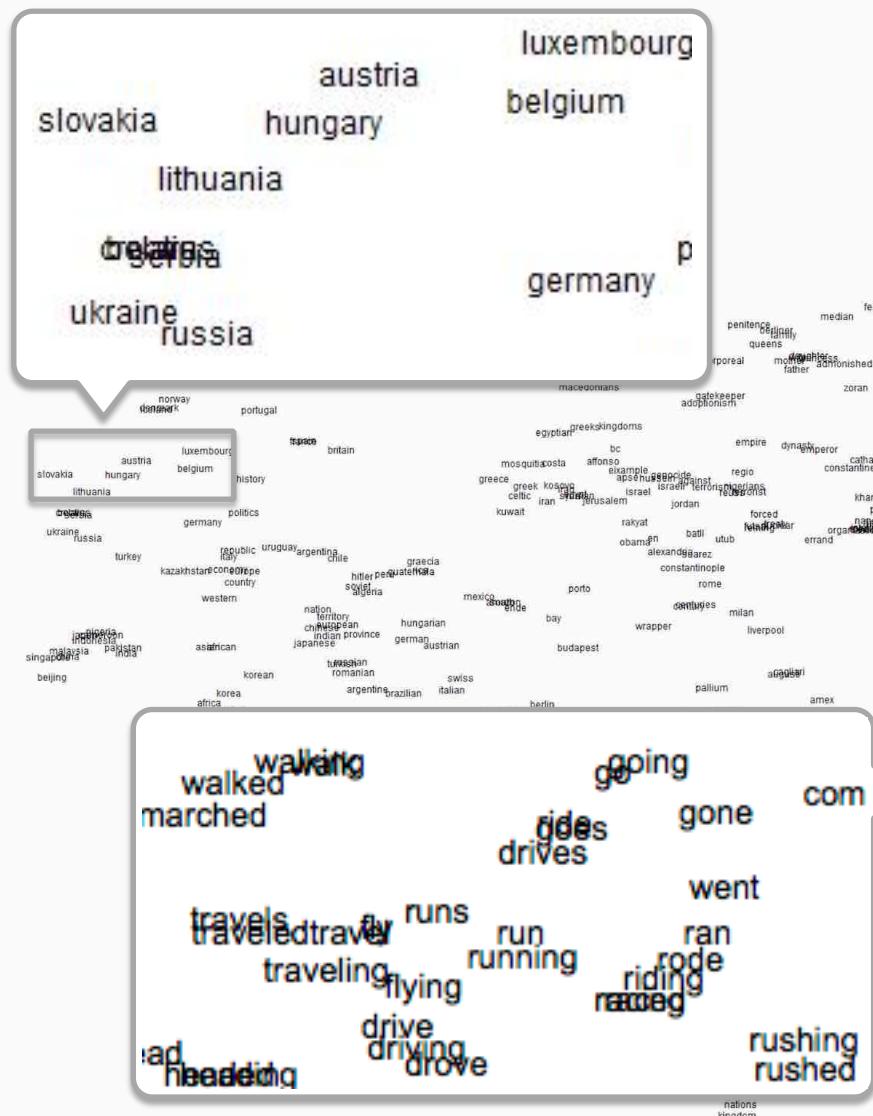
Word Embedding Model



- When vector representations are dense, they are often called **embedding** e.g. word embedding

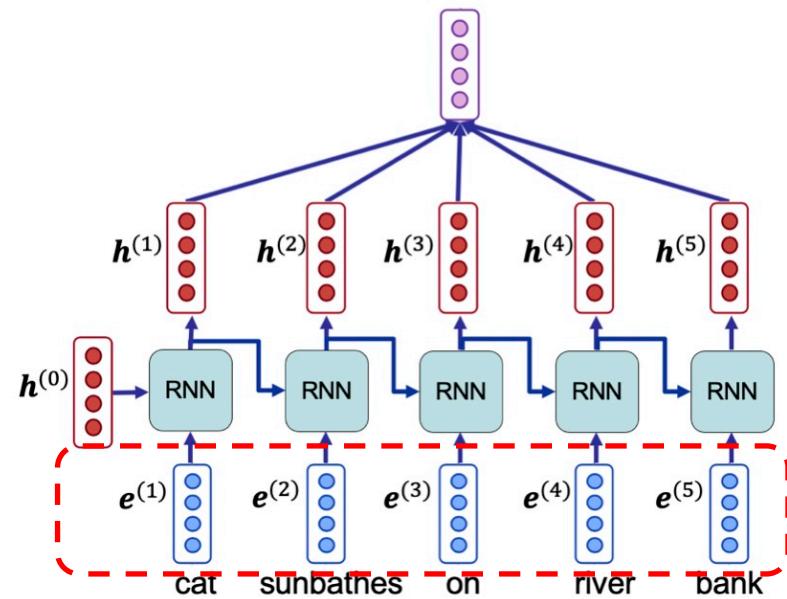
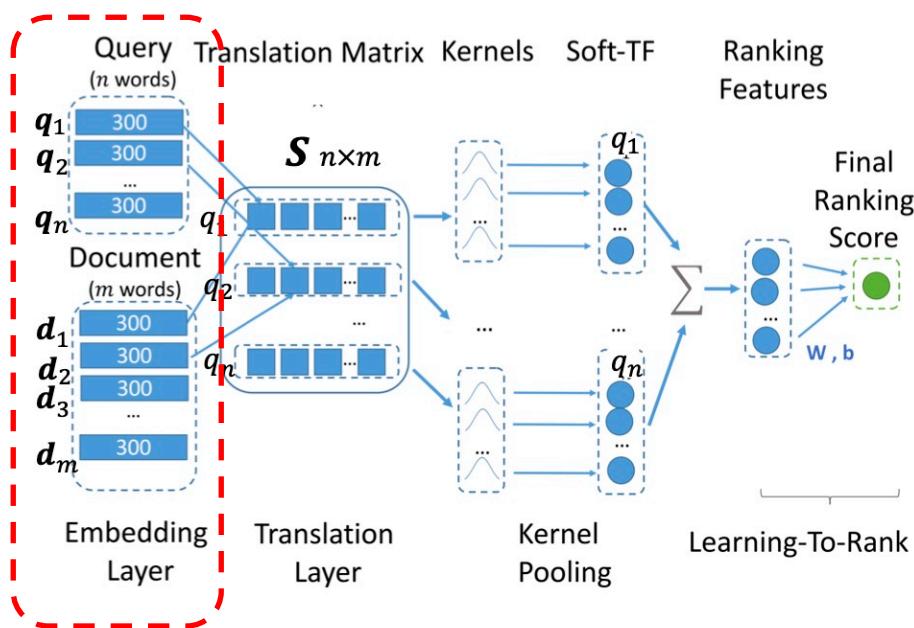
Word Embedding – Visualization

T-SNE



Word embeddings projected to a two-dimensional space

Word Embedding – Building Block of Modern NLP



Distributional Semantics

“A word’s **meaning** is given by the words that frequently appear close-by”

- The **context** of a word w is the set of words appear in the nearby of w (e.g. within a fixed-size window)

One of the most successful ideas of modern statistical NLP!



drink

sacred

beer

Tesgüino

ritual

fermented

Mexico

corn

Tarahumara people

lager

dark brown

Märzen

brew

malt

Bavaria

bar

drink

beer



drink

sacred

beer

ritual

brew

corn

Tesgüino

Mexico

Tarahumara people





lager

dark brown

bar

Märzen

beer

bottle

brew

malt

Bavaria

drink

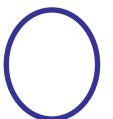
Tesgüino ←→ **Märzen**



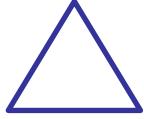
Two words are **related** when they have many common context-words

Märzen

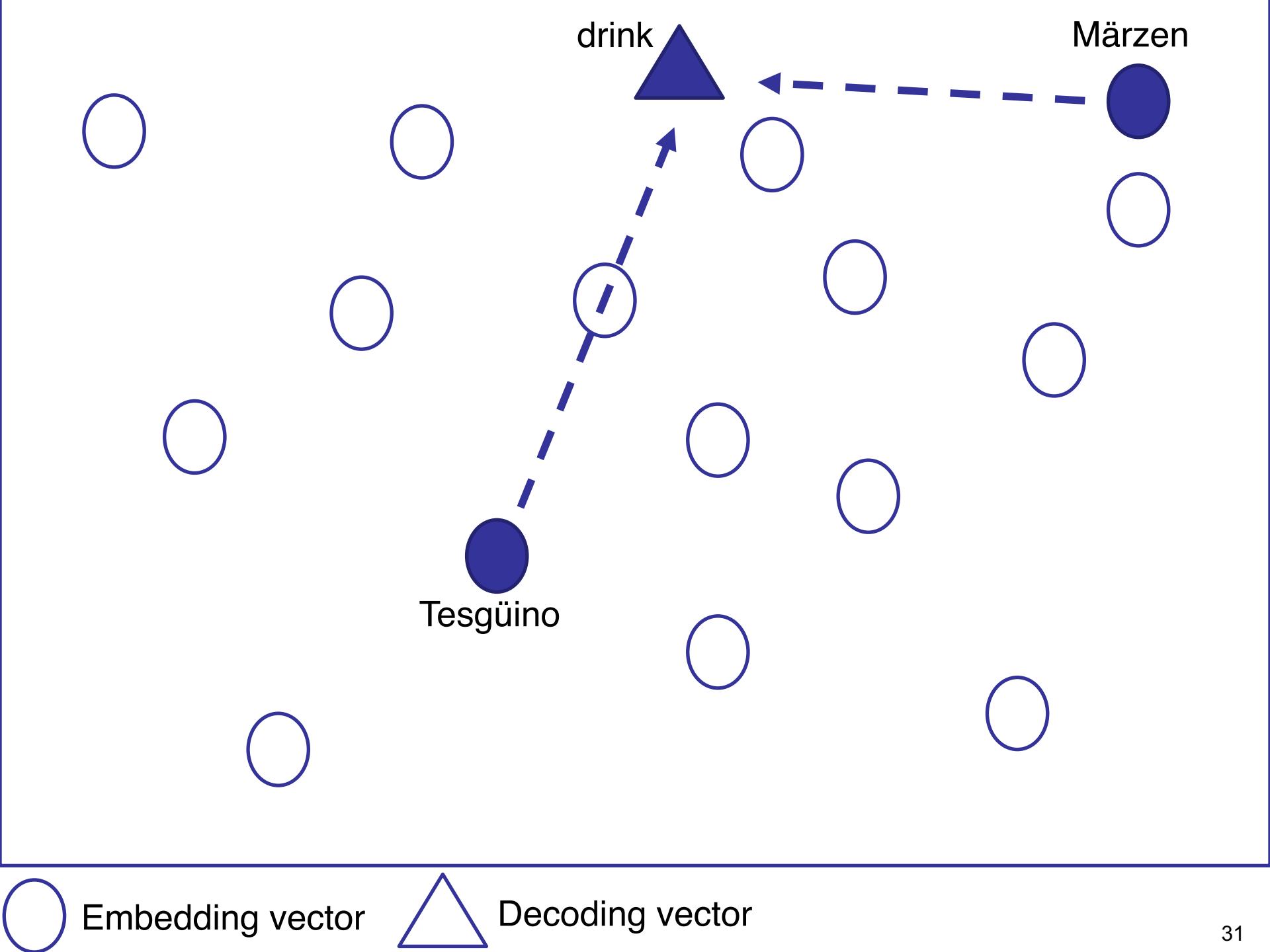
Tesgüino



Embedding vector

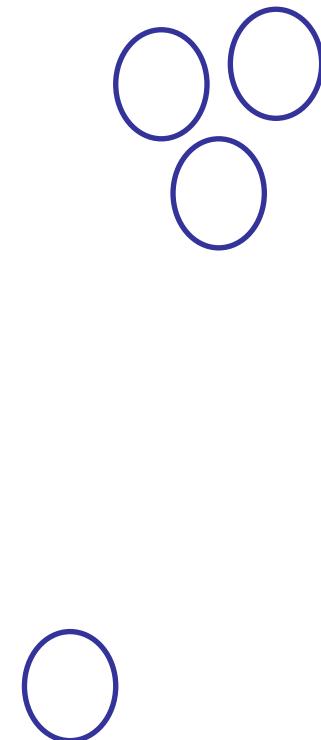
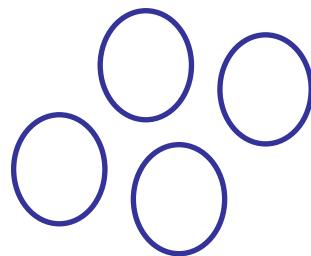
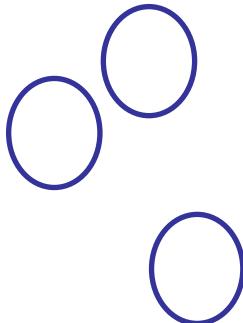


Decoding vector



drink Märzen

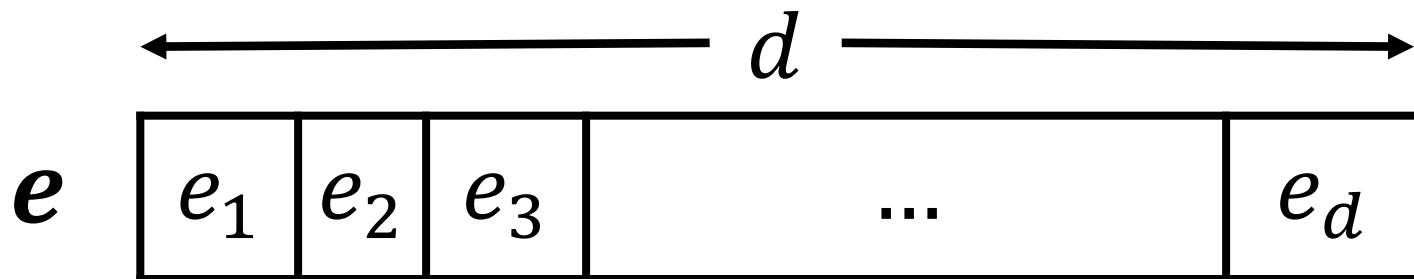
Tesgüino

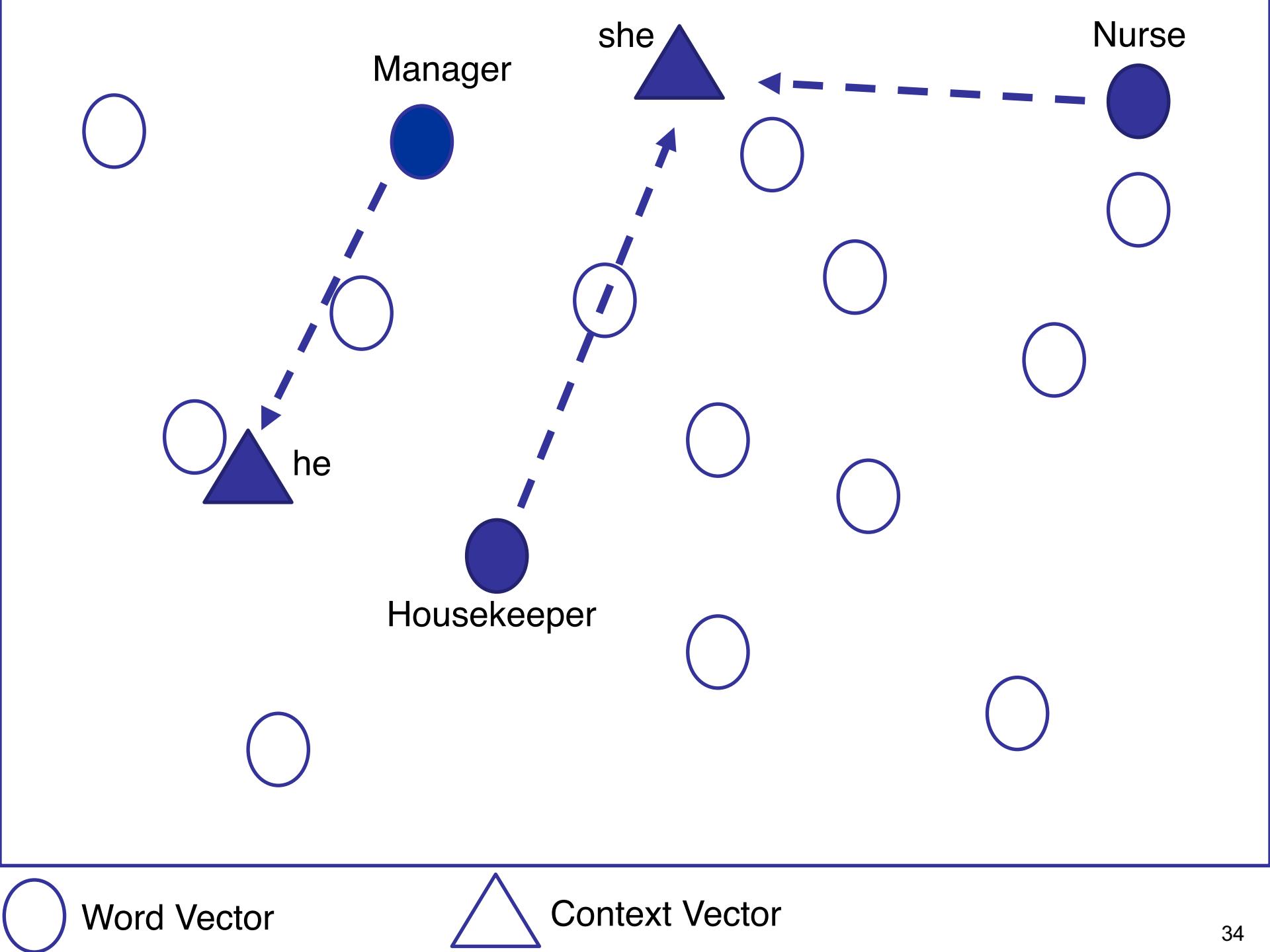


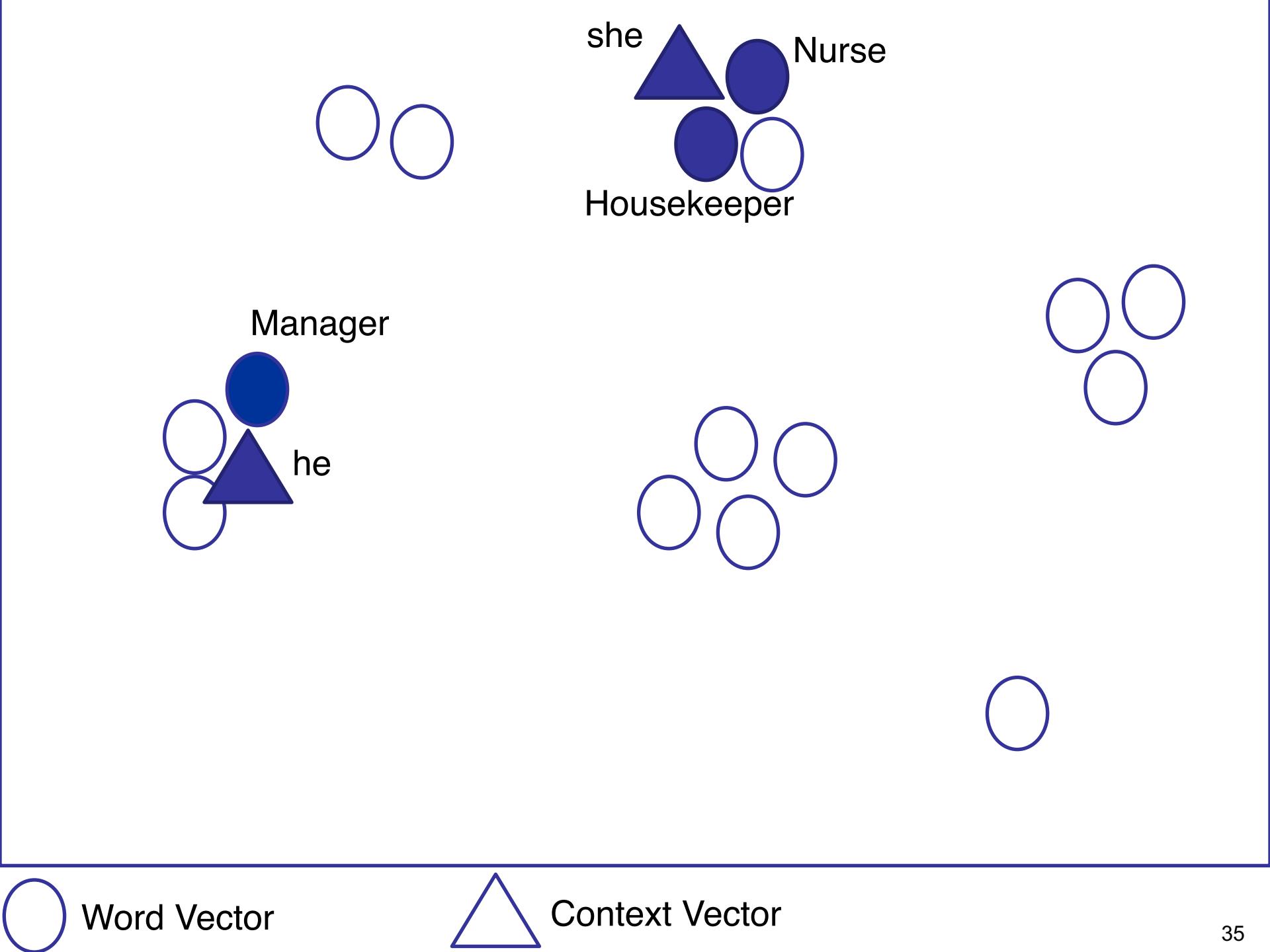
○ Embedding vector ▲ Decoding vector

Representation learning and bias

Representation learning encodes information but also may encode the **underlying biases** in data!







Bias in word analogies

- Recap – word analogy: **man** to **woman** is like **king** to ? (**queen**)

$$\begin{aligned}x_{\text{king}} - x_{\text{man}} + x_{\text{woman}} &= x^* \\x^* &\approx x_{\text{queen}}\end{aligned}$$

- Gender bias is reflected in word analogies

Gender stereotype *she-he* analogies

sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant

Gender appropriate *she-he* analogies

queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

How to measure bias using word embeddings?

- Bias: discrepancy between two concepts \mathbb{Z} and $\tilde{\mathbb{Z}}$, e.g. between female and male*
- A (second-order) bias measurement method:

$$\text{BIAS}_{\text{2ND}}(w) = \frac{1}{|\mathbb{Z}|} \sum_{z \in \mathbb{Z}} \cos(\mathbf{e}_z, \mathbf{e}_w) - \frac{1}{|\tilde{\mathbb{Z}}|} \sum_{\tilde{z} \in \tilde{\mathbb{Z}}} \cos(\mathbf{e}_{\tilde{z}}, \mathbf{e}_w)$$

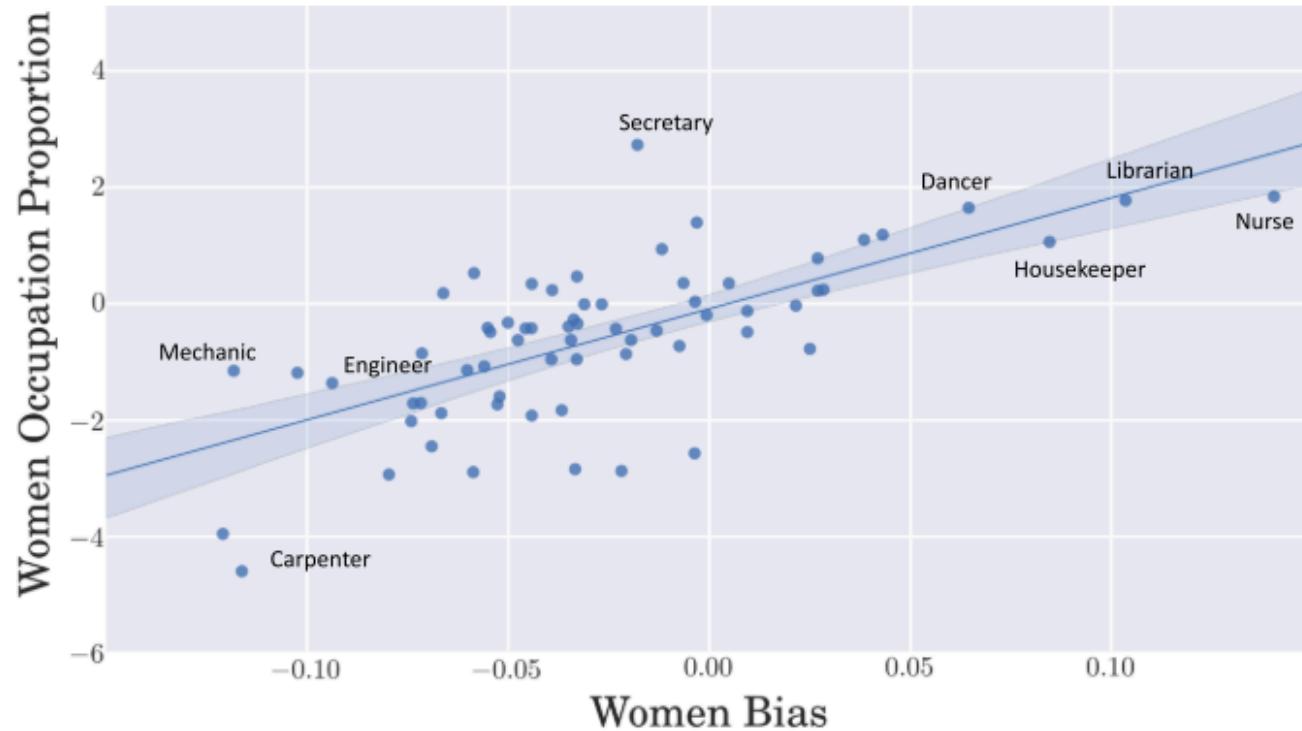
- \mathbf{e}_w is the vector of word w in a pre-trained word embedding (such as word2vec or GloVe)
- Sample \mathbb{Z} and $\tilde{\mathbb{Z}}$ when measuring bias towards **female**:

$$\mathbb{Z} = \{\mathbf{she}, \mathbf{her}, \mathbf{woman}, \mathbf{girl}\}$$

$$\tilde{\mathbb{Z}} = \{\mathbf{he}, \mathbf{him}, \mathbf{man}, \mathbf{boy}\}$$

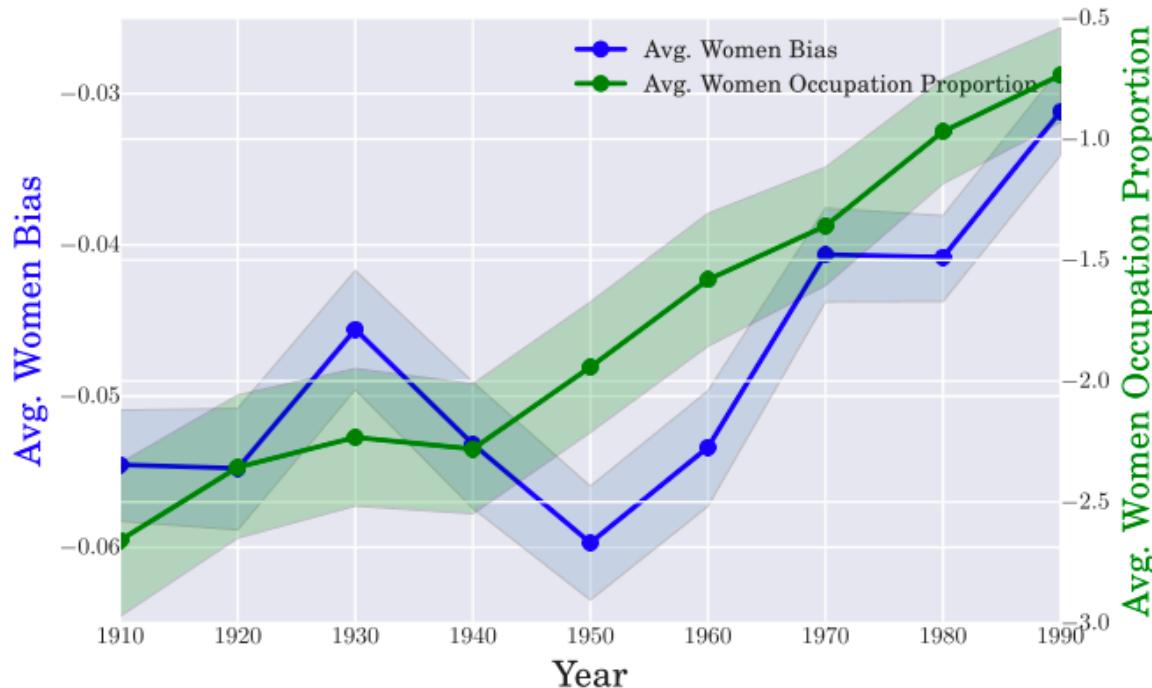
Defining *gender* as a binary construct – namely female vs. male – is an *unpleasant* simplification, as it neglects the wide definition of gender!

Word Embeddings capture societal realities!



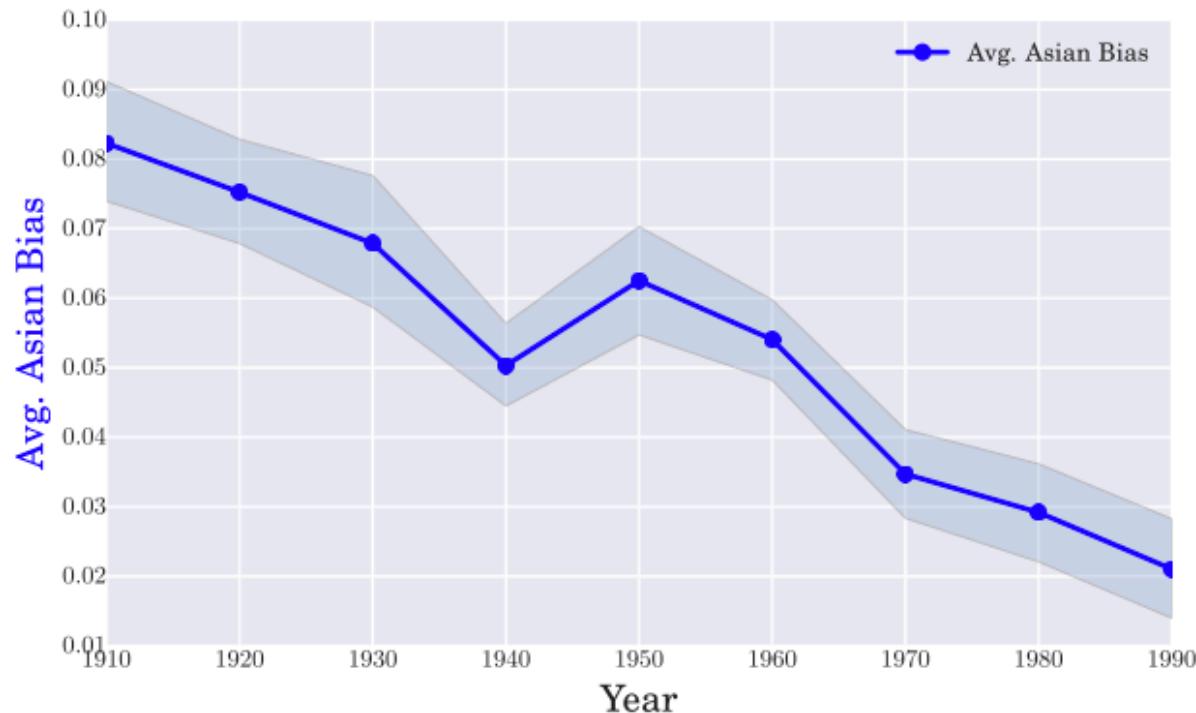
(a) Woman occupation proportion vs embedding bias in Google News vectors. More positive indicates more women biased on both axes. $p < 10^{-9}$, r-squared = .462.

Word Embeddings capture societal realities!



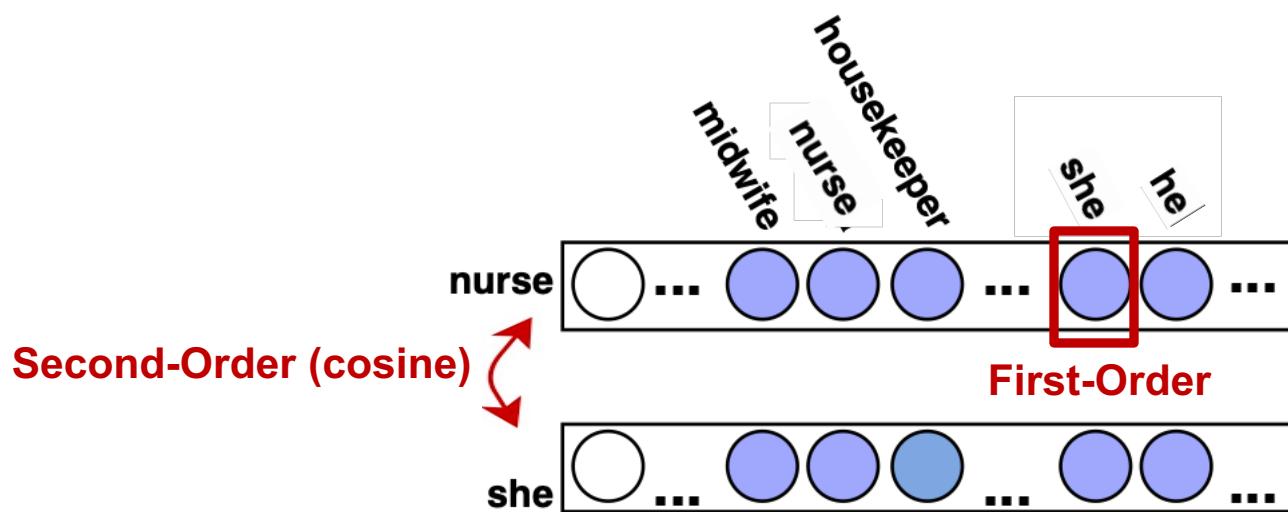
(b) Average gender bias score over time in COHA embeddings in occupations vs the average log proportion. In blue is relative women bias in the embeddings, and in green is the average log proportion of women in the same occupations.

Word Embeddings capture societal realities!



(c) Asian bias score over time for words related to the outsiders in COHA data.

Bias Measurement Methods

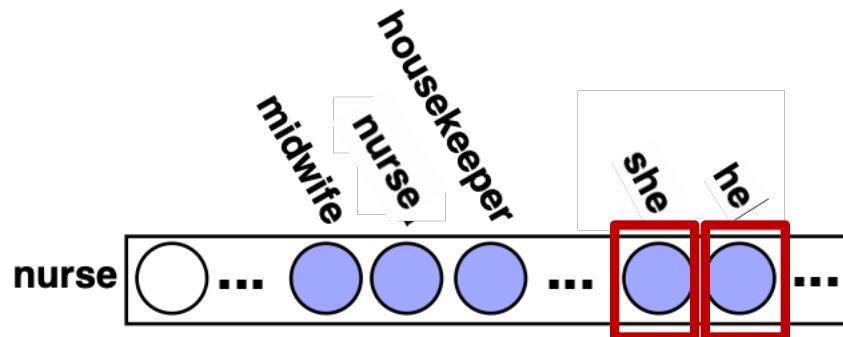


- **First-Order co-occurrence:** co-occurrence of two words in the same context
 - Examples: **drink** to **beer** or **drink** to **wine**
- **Second-Order co-occurrence:** similarity of two word vectors → usually by using cosine
 - Examples: **beer** to **wine**, or **Tesgüino** to **Märzen**

Bias measurement using word embeddings

- Recap: Second-order bias measurement for word w :

$$\text{BIAS}_{\text{2ND}}(w) = \frac{1}{|\mathbb{Z}|} \sum_{z \in \mathbb{Z}} \cos(e_z, e_w) - \frac{1}{|\tilde{\mathbb{Z}}|} \sum_{\tilde{z} \in \tilde{\mathbb{Z}}} \cos(e_{\tilde{z}}, e_w)$$



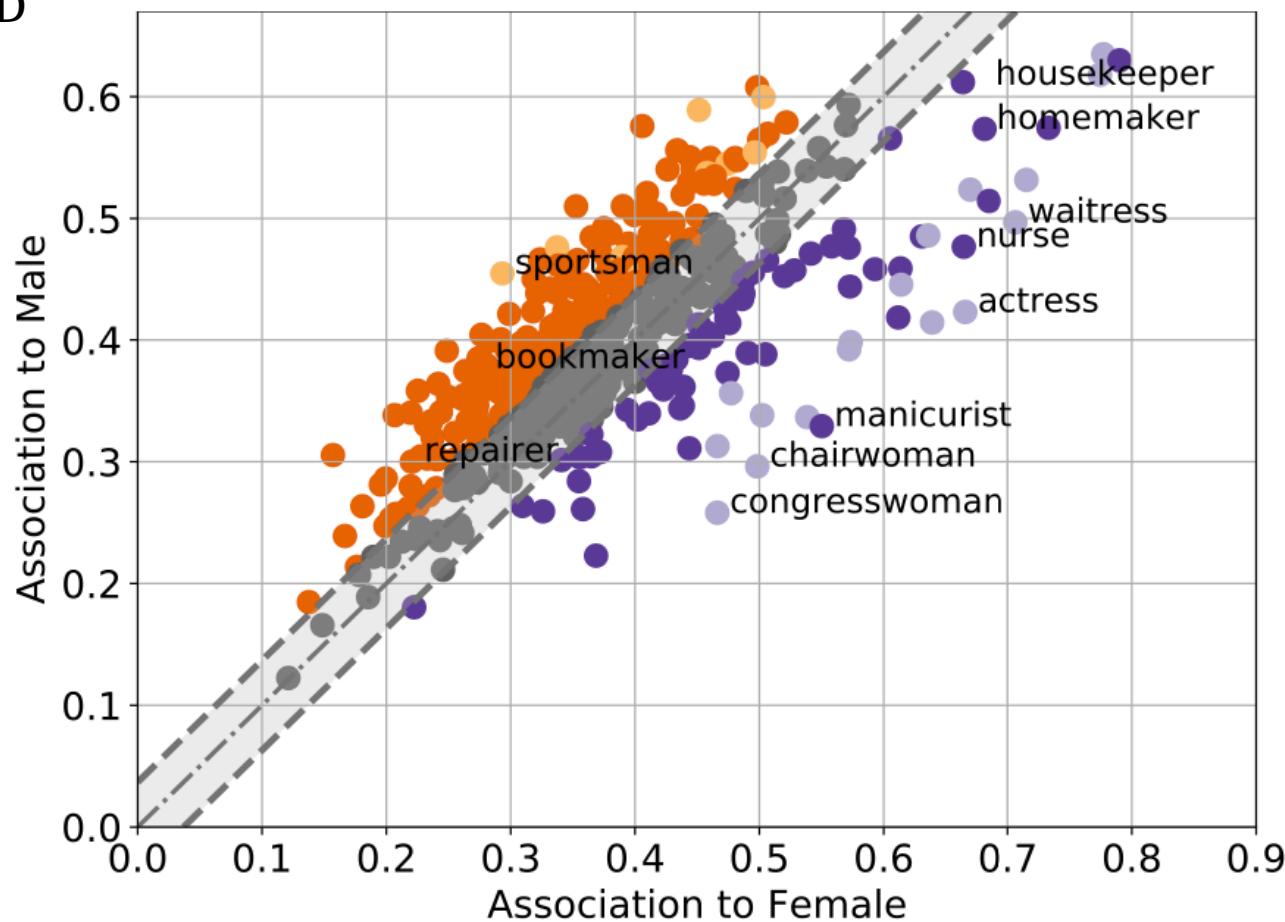
- First-order bias measurement for word w :

$$\text{BIAS}_{\text{1ST}}(w) = \frac{1}{|\mathbb{Z}|} \sum_{z \in \mathbb{Z}} f(z, w) - \frac{1}{|\tilde{\mathbb{Z}}|} \sum_{\tilde{z} \in \tilde{\mathbb{Z}}} f(\tilde{z}, w)$$

f denotes a measured first-order co-occurrence between the words

Measuring societal biases with second-order co-occurrences

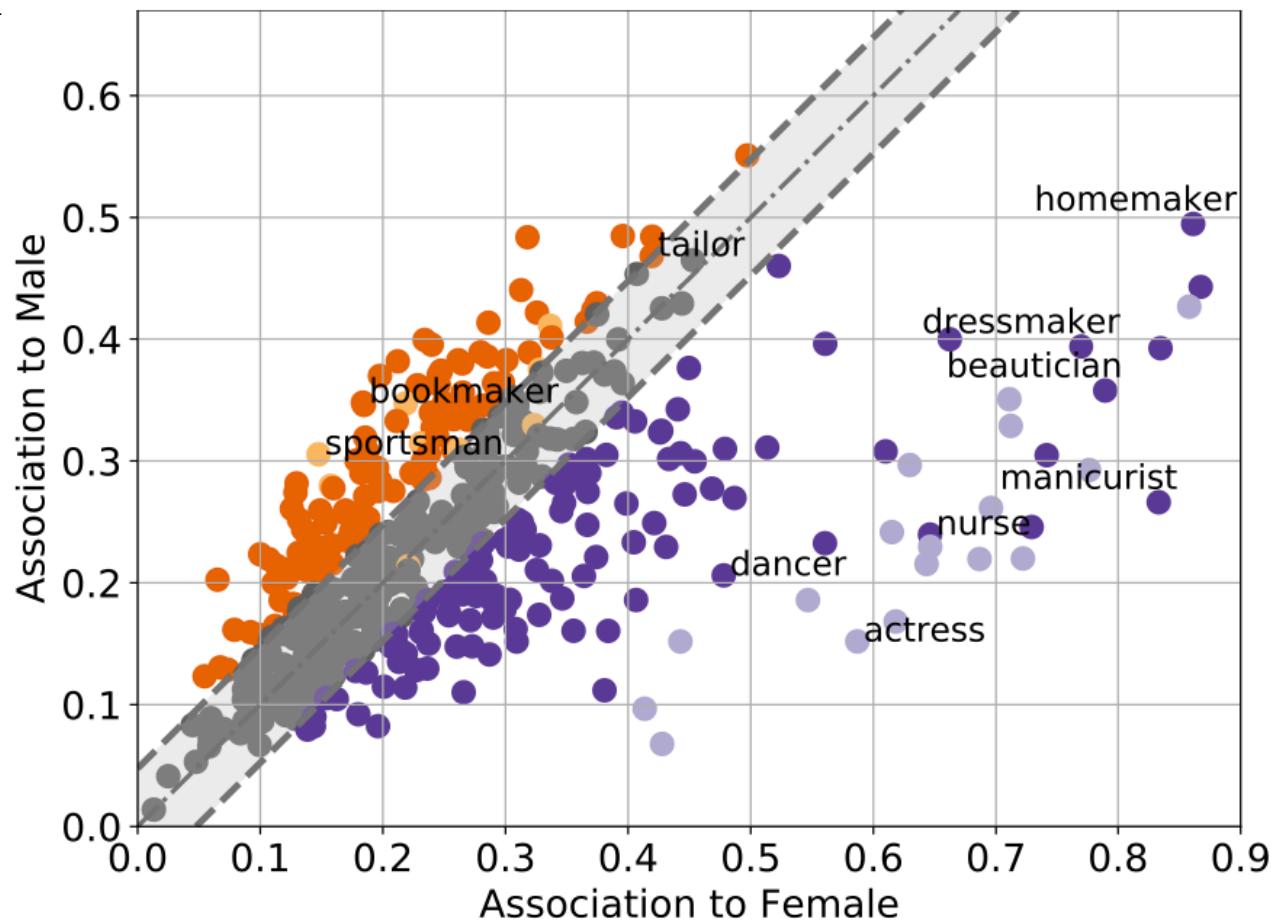
BIAS_{2ND}



Associations are measured using a word2vec model, trained on a recent Wikipedia corpus

Measuring societal biases with first-order co-occurrences

BIAS_{1ST}



Associations are measured using a word2vec model, trained on a recent Wikipedia corpus

Rekabsaz N., Henderson J., West R., and Hanbury A. "Measuring Societal Biases in Text Corpora via First-Order Co-occurrence." *arXiv preprint arXiv:1812.10424* (2020).

Correlations with job market statistics

Order	Representation	Method	Labor Data		Census Data	
			Spearman ρ	Pearson's r	Spearman ρ	Pearson's r
High-Order	PMI	DIRECTIONAL	0.28	0.07	0.18	0.02
		CENTROID	0.14	0.21	0.35	0.40
		AVERAGE _{HIGH}	0.33	0.24	0.27	0.19
	PMI-SVD	DIRECTIONAL	0.05	0.07	0.00	0.00
		CENTROID	0.41	0.47	0.46	0.53
		AVERAGE _{HIGH}	0.41	0.49	0.49	0.56
First-Order	PMI	AVERAGE _{FIRST}	0.53	0.51	0.57	0.62
High-Order	PPMI	DIRECTIONAL	0.45	0.49	0.39	0.47
		CENTROID	0.43	0.46	0.45	0.50
		AVERAGE _{HIGH}	0.43	0.46	0.45	0.52
	PPMI-SVD	DIRECTIONAL	0.05	0.07	0.00	0.00
		CENTROID	0.41	0.47	0.46	0.53
		AVERAGE _{HIGH}	0.41	0.49	0.49	0.56
First-Order	PPMI	AVERAGE _{FIRST}	0.59	0.58	0.64	0.64
High-Order	SPPMI	DIRECTIONAL	0.26	0.37	0.26	0.28
		CENTROID	0.39	0.45	0.45	0.48
		AVERAGE _{HIGH}	0.32	0.40	0.44	0.48
	SPPMI-SVD	DIRECTIONAL	0.17	0.29	0.11	0.03
		CENTROID	0.28	0.35	0.39	0.43
		AVERAGE _{HIGH}	0.26	0.38	0.36	0.46
First-Order	SPPMI	AVERAGE _{FIRST}	0.57	0.49	0.52	0.48
High-Order	GloVe	DIRECTIONAL	0.53	0.56	0.34	0.46
		CENTROID	0.58	0.60	0.39	0.51
		AVERAGE _{HIGH}	0.60	0.60	0.39	0.51
First-Order	initGlove eGloVe	AVERAGE _{FIRST}	0.38	0.42	0.40	0.51
High-Order	SG	DIRECTIONAL	0.50	0.54	0.58	0.64
		CENTROID	0.55	0.57	0.60	0.65
		AVERAGE _{HIGH}	0.55	0.57	0.59	0.65
First-Order	eSG	AVERAGE _{FIRST}	0.66	0.61	0.67	0.70

Correlation results of the gender bias values, calculated with word representations, to the statistics of the portion of women in occupations

Summary

What we know so far ...

- Word embeddings capture and **encode societal biases**, reflected in the underlying corpora
 - These biases also exist in contextualized word embeddings
- Word embeddings enable the study of **societal phenomena**
 - e.g. monitoring how gender/ethnicity/etc. is perceived during time

Subsequent question:

- What about bias in down-stream NLP tasks?
 - Existence of bias could become problematic in many NLP tasks such as *job search, content-based recommendation systems, IR, sentiment analysis, etc.*
- Since the pre-trained word embeddings are widely used in NLP tasks, are biases in word embeddings also transferred to the tasks?

Agenda

- Societal biases in NLP ... what? why?
- Bias in word embedding
- **Measuring bias in Information Retrieval**

Gender bias in IR

- Previous studies
 - Kay et al. [2015], Otterbacher et al. [2017], Chen et al. [2018], Biega et al. [2018], Otterbacher et al. [2018], ...
- How to measure gender bias, especially according to the content of retrieved documents?

Novel collection for gender bias measurement in IR

- What is the effect of collection / ranking models / transfer learning on gender bias?

Experiments on various recent (neural) IR models

Step1: Gendered vs. non-gendered queries annotation

- Sets of queries, annotated by Amazon Turkers according to their gender-related contents
 - Queries are taken from MS MARCO dev set
- *Non-gendered queries*: queries with no indication of gender

Gendered	<i>Non-gendered</i> (1765)	what is a synonym for beautiful what is the meaning of resurrect
	<i>Female</i> (742)	who was oprah winfrey earliest pregnancy symptoms
	<i>Male</i> (1202)	where is martin luther king jr's place who was the king of ancient rome
	<i>Other or Multiple Genders</i> (41)	is g dragon gay how long was shakespeare married to anne

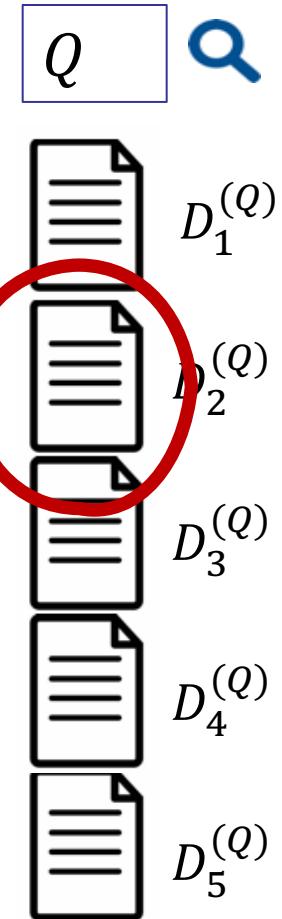
Step 2: document female/male magnitude

- In what extent each document D contains female/male topics?
- $\gamma^{\mathbb{Z}}(D)$: the **magnitude** of concept \mathbb{Z} in document D

$$\gamma^{\mathbb{Z}}(D) = \sum_{t \in \mathbb{Z}} \log \text{tc}_{t,D}$$

$$\gamma^{\tilde{\mathbb{Z}}}(D) = \sum_{t \in \tilde{\mathbb{Z}}} \log \text{tc}_{t,D}$$

- \mathbb{Z} Male definitional words: ***he, him, man, boy***, etc.
- $\tilde{\mathbb{Z}}$ Female definitional words : ***she, her, woman, girl***, etc.



* *Limitation: gender is conceived as a binary construct!*

Step 3: ranking bias metric

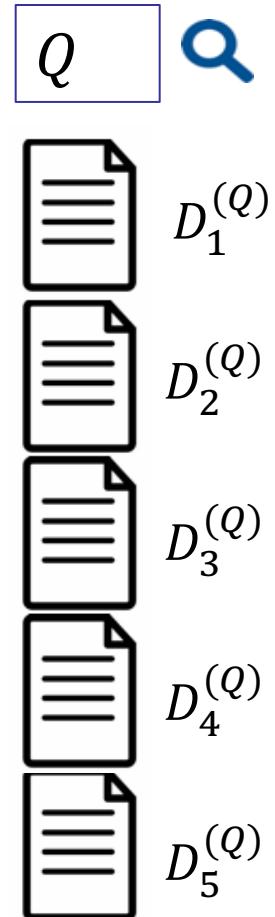
- Rank Bias (RaB) metric
 - For the given query Q :

$$q\text{RaB}(Q) = \frac{1}{t} \sum_{i=1}^t \gamma^{\mathbb{Z}}(D_i^{(Q)}) - \gamma^{\tilde{\mathbb{Z}}}(D_i^{(Q)})$$

- RaB of a model is the average over the queries \mathbb{Q} :

$$\text{RaB} = \frac{1}{|\mathbb{Q}|} \sum_{Q \in \mathbb{Q}} q\text{RaB}(Q)$$

- Another metric: Average Ranking Bias (ARaB)

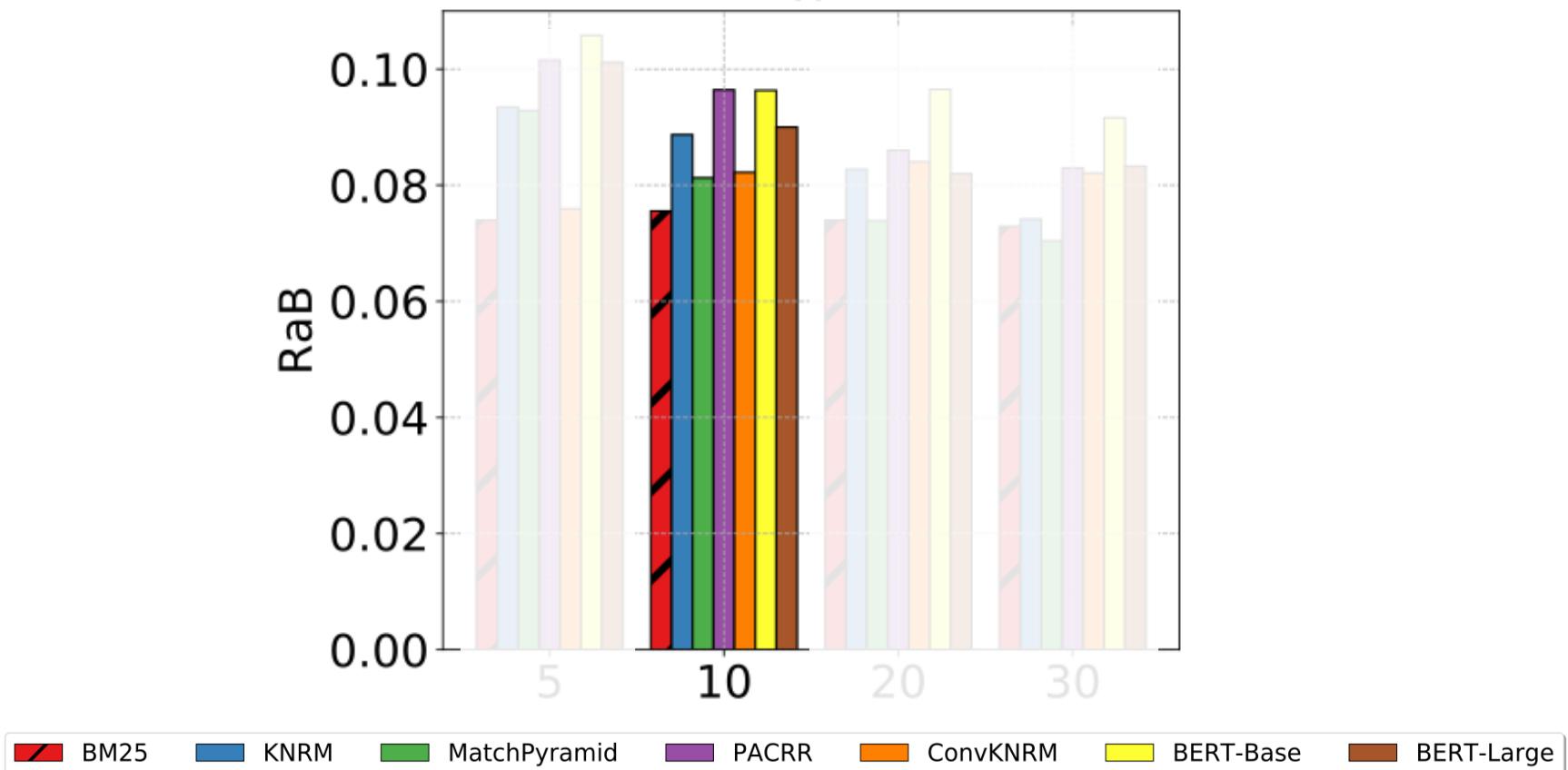


Experiment design

- Ranking models
 - BM25
 - Various neural ranking models
 - Word embeddings are initialized with GloVe or Randomly (*RND*)
 - BERT models
- Evaluated on MS-MARCO passage retrieval

Ranking Model	Model Parameters		Evaluation	
	All	Transferred	MRR	Recall
BM25		-	0.192	0.398
KNRM _{RND}	109,481,411	none	0.213	0.390
KNRM		GloVe	0.230	0.439
MatchPyramid _{RND}	109,539,960	none	0.232	0.424
MatchPyramid		GloVe	0.240	0.445
PACRR _{RND}	109,875,938	none	0.228	0.426
PACRR		GloVe	0.242	0.451
ConvKNRM _{RND}	110,022,399	none	0.243	0.443
ConvKNRM		GloVe	0.268	0.488
BERT-Base	109,483,778	all	0.342	0.585
BERT-Large	335,143,938	all	0.353	0.596

Results



- All models show an overall bias towards male
- Neural models show higher gender bias in comparison with BM25!
 - Especially, fine-tuned BERT models show higher bias than other neural models

Effect of transfer learning

	Cut-off: 10			
BM25	0.076	0.076	0.039	0.031
KNRM	0.089 ($\downarrow 0.020$)	0.090 ($\downarrow 0.004$)	0.041 ($\downarrow 0.007$)	0.037 ($\uparrow 0.003$)
MatchPyramid	0.081 ($\downarrow 0.002$)	0.089 ($\downarrow 0.006$)	0.047 ($\downarrow 0.007$)	0.047 ($\downarrow 0.007$)
PACRR	0.096 ($\downarrow 0.008$)	0.104 ($\downarrow 0.019$)	0.043 ($\uparrow 0.002$)	0.042 ($\downarrow 0.001$)
ConvKNRM	0.082 ($\downarrow 0.001$)	0.066 ($\uparrow 0.009$)	0.045 ($\downarrow 0.005$)	0.037 ($\downarrow 0.004$)
BERT-Base	0.096	0.102	0.056	0.057
BERT-Large	0.090	0.100	0.052	0.054

- Arrows show the changes in RaB, when word embeddings are initialized randomly instead of initialization with a pre-trained GloVe
- Models without transfer learning show smaller degrees of gender bias → Transfer learning increases gender bias!

Bias in IR – Summary

- We see a framework to measure gender bias in the content of the retrieved results
- Neural ranking models, especially BERT models, increase gender bias
- Transfer learning increases bias

Code and resources: https://github.com/navid-rekabsaz/GenderBias_IR

Final words: About “debiasing”

- **Debiasing:** methods to **mitigate** or **reduce bias**
 - The aim is to make the output or decision of a model **agnostic** to **sensitive features** (such as gender, race, ethnicity, age)
- Debiasing approaches in NLP literature address ...
 - **Datasets / Corpora:** by changing/adding/removing data in collection
 - **Models**
 - By adding debiasing/fairness criteria to model's objective function
 - By training adversarial networks to remove sensitive information in learned representations
 - **Output results:** by post-processing model's outputs

Questions?

