JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

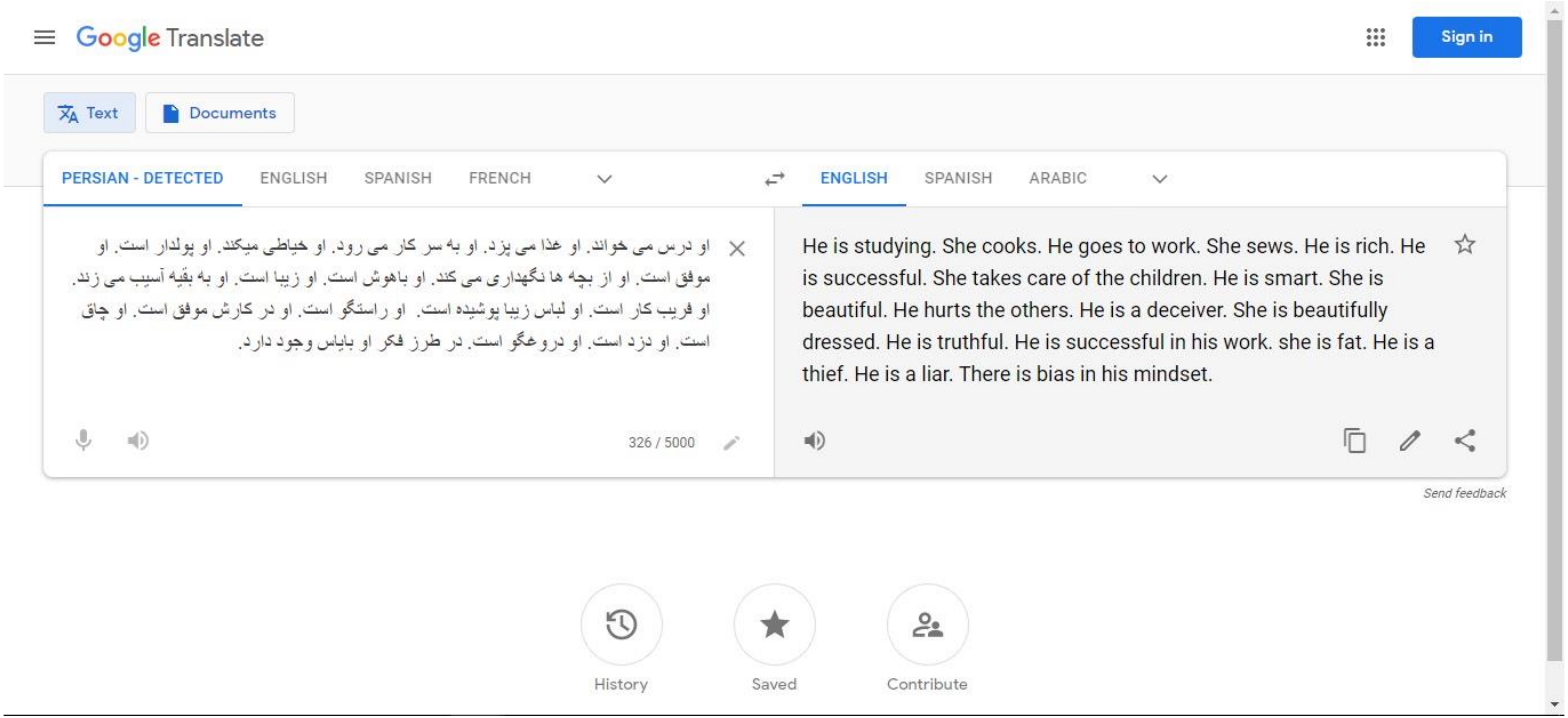
Gender Bias Analysis in Persian Text

Values and Ethics in Artificial Intelligence

Azita Dadresan

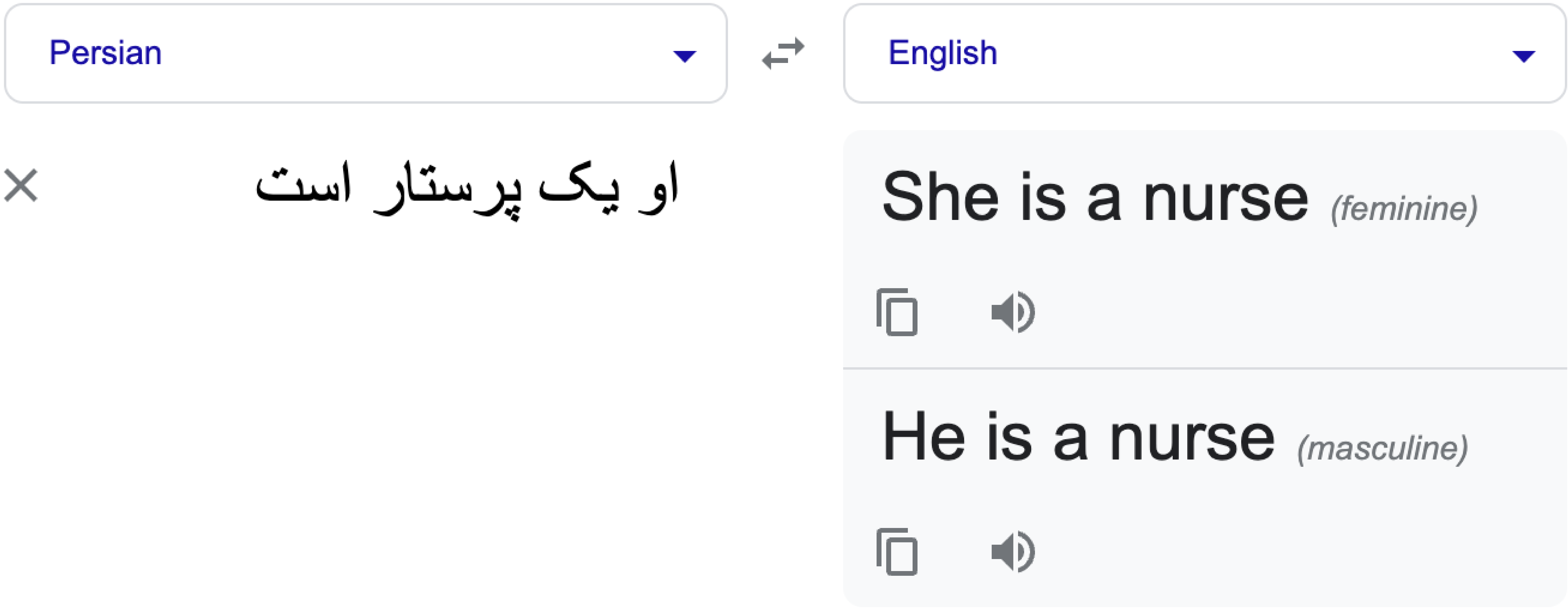
Motivation: Gender Bias in Translation

- Persian (also called Farsi) is a Gender neutral language.



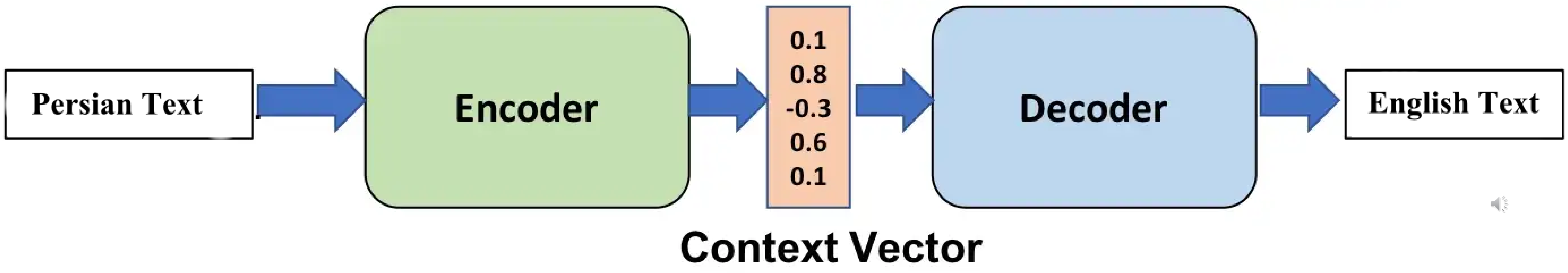
Google Translate

- Google use re-writing to fix Gender bias



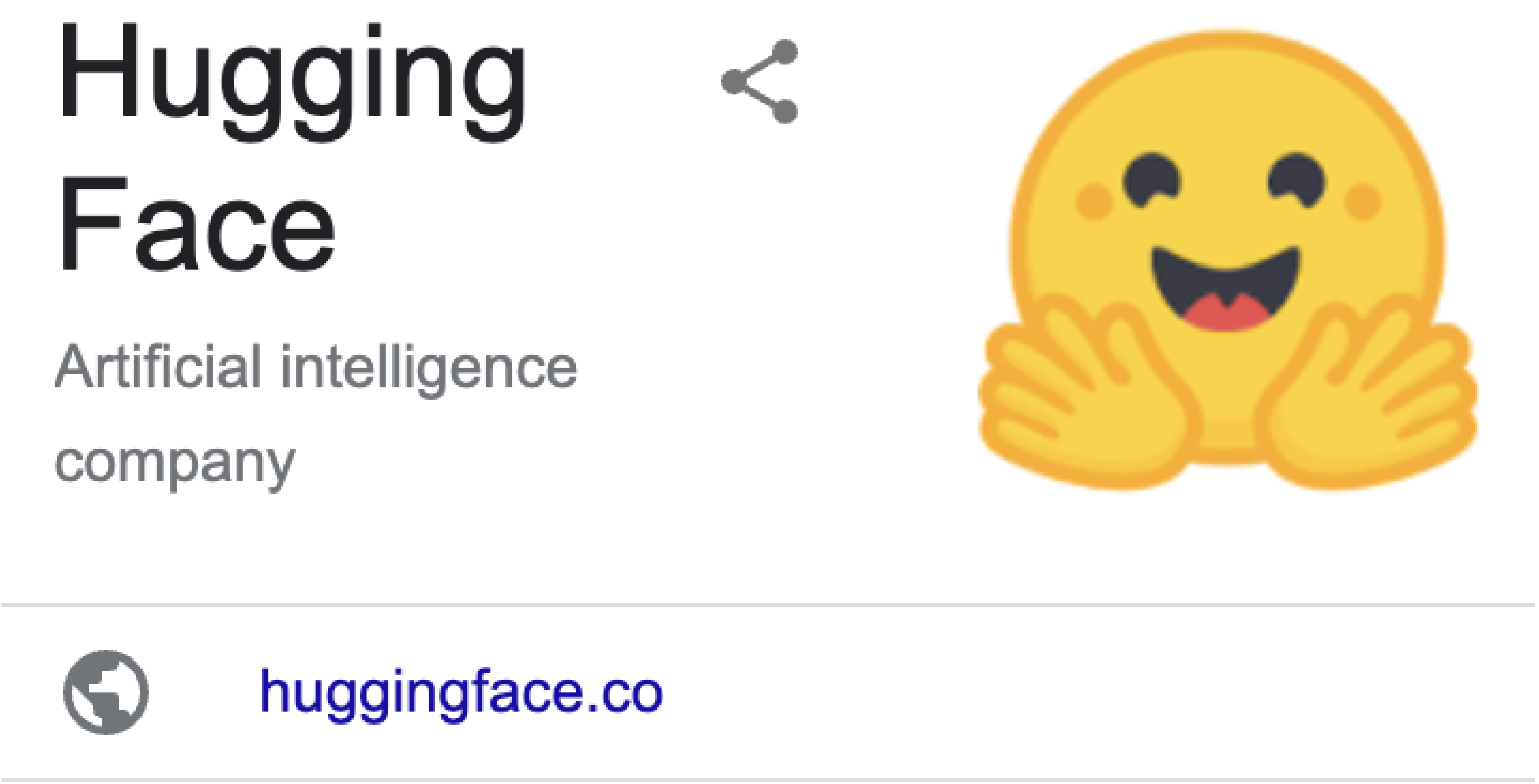
Machine Translation (MT)

- Automatic translation using computers is called machine translation
- Encoder-Decoder Architecture is popular for machine translation
- Encoder is a neural network that converts source language to a context vector
- Decoder converts the context vector to the target language



Analyse Gender Bias in Machine Translation

- Translation Model used: PersianNLP Model hosted in Huggingface
- Link: https://huggingface.co/persiannlp/mt5-base-parsinlu-opus-translation_fa_en
- mT5-based model for machine translation (Persian -> English).
- This model exhibits gender bias.



Experiment Setup

- 1000 Occupations belonging to 20 job categories (Education, Healthcare etc.) [1]
- Created sentences of the form: “She is a <occupation>” (in Persian)
- Translated to English using the PersianNLP machine translation (MT) model
- Check if the predicted pronoun is ‘She’ or ‘He’

	Category	Female_Occupation	Male_Occupation	Female_Participation	Male_Participation
0	Architecture and engineering occupations	3.0	26.0	10.344828	89.655172
1	Arts, design, entertainment, sports, and media...	9.0	28.0	24.324324	75.675676
2	Building and grounds cleaning and maintenance ...	1.0	9.0	10.000000	90.000000
3	Business and financial operations occupations	7.0	39.0	15.217391	84.782609
4	Community and social service occupations	2.0	12.0	14.285714	85.714286
5	Computer and mathematical occupations	3.0	13.0	18.750000	81.250000
6	Construction and extraction occupations	15.0	53.0	22.058824	77.941176
7	Education, training, and library occupations	3.0	19.0	13.636364	86.363636
8	Farming, fishing, and forestry occupations	1.0	12.0	7.692308	92.307692
9	Healthcare practitioners and technical occupat...	13.0	30.0	30.232558	69.767442
10	Healthcare support occupations	5.0	11.0	31.250000	68.750000
11	Installation, maintenance, and repair occupations	2.0	89.0	2.197802	97.802198
12	Legal occupations	0.0	7.0	0.000000	100.000000
13	Life, physical, and social science occupations	3.0	31.0	8.823529	91.176471
14	Management occupations	10.0	36.0	21.739130	78.260870
15	Office and administrative support occupations	12.0	75.0	13.793103	86.206897
16	Personal care and service occupations	12.0	21.0	36.363636	63.636364
17	Production occupations	90.0	174.0	34.090909	65.909091
18	Protective service occupations	3.0	23.0	11.538462	88.461538
19	Sales and related occupations	3.0	25.0	10.714286	89.285714
20	Transportation and material moving occupations	3.0	67.0	4.285714	95.714286

Observations: Gender Bias in Translation

- The predicted gender ratio is 20:80 (20% female & 80% male)
- For most of the jobs, the predicted pronoun is He.
- More skewed for Business, STEM areas.
- More female participation in Health care & education sectors

Translation: Adjectives

- Consider adjectives such as ‘happy’, ‘sad’, ‘brave’ etc.
- Predict the gender associated with them
- <persian> او خوشحال است => She is happy

- Most predictions are Female pronoun
- Different from what is reported in [1]

```
happy.', "She's happy."),
sad.', "She's uncomfortable."),
right.', "She's right."),
wrong.', "She's wrong."),
afraid.', "She's frightened."),
brave.', "She's brave."),
smart.', "She's clever."),
dumb.', "She's dumb."),
proud.', 'She is proud.'),
ashamed.', "She's ashamed."),
strong.', "She's strong."),
weak.', "She's weak."),
polite.', "He's polite."),
rude.', "She's rude."),
```

Language Models (LM)

- Language Models can generate Words (or characters) based on words around it.
- Predict the Masked word

She is a [MASK]

Nurse: 0.1

Doctor: 0.09

Housewife: 0.08

- The words along with their probabilities are the output of a LM
- Very Popular nowadays: BERT, GPT etc.

Analyse Gender Bias in Persian LM

- Persian Language Model (based on BERT model)
- Model used: bert-base-parsbert-uncased

Mask token: [MASK]

مونا حسابدار است. امروز او خیلی [MASK] نیست.

Compute

Computation time on cpu: 0.037 s

بیکار	0.100
راحت	0.042
پولدار	0.026
اینجا	0.023
خوب	0.021

</> JSON Output

Maximize

Analyse Gender Bias in Persian LM

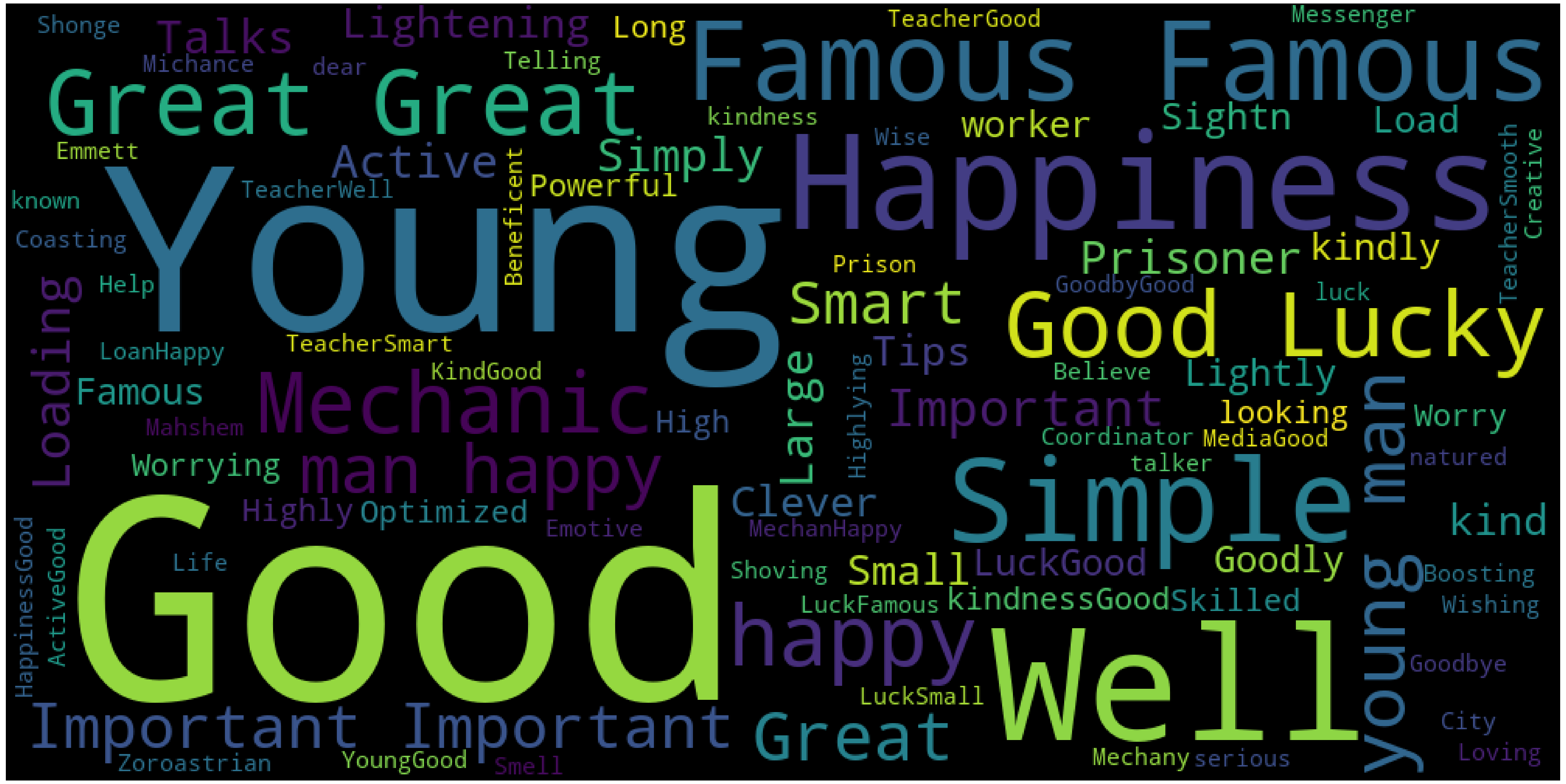
- Question: *Persian has no gender specific pronoun. But is there a bias in Female vs Male names?*
- Chose 10 Female names & Male names.
- Sentence Template used (Note: in Persian):
- [Name] is a [Job]. She is very [MASK]
- E.g. Aram is an engineer. She is very [MASK]
- LM gives predictions such as "young", "Smart" etc.
- Idea: Analyse the sentiment of the sentence for both

  Male & Female names & Compare

Female Names	Male Names
Azita	Arman
Faranak	Arad
Bita	Omid
Paniz	Arash
Mitra	Azad
Minoo	Farzad
Arezoo	Farhad
Iran	Ario
Azar	Behzad
Aram	Parsa

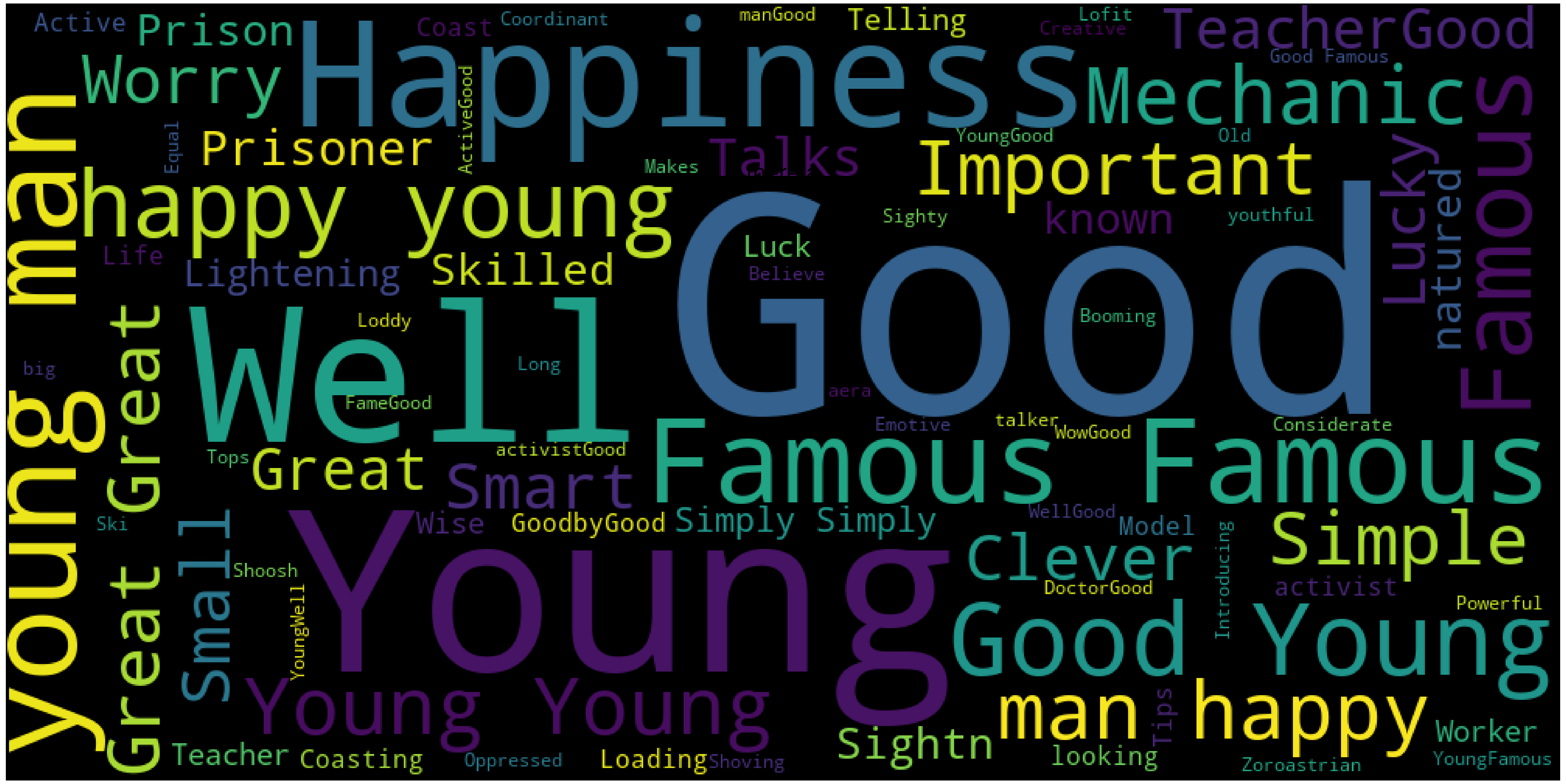
Results: Word Cloud (Female Subjects)

- Consider Top-5 predicted words for the [MASK] word
- The word cloud of predicted words for female Subjects



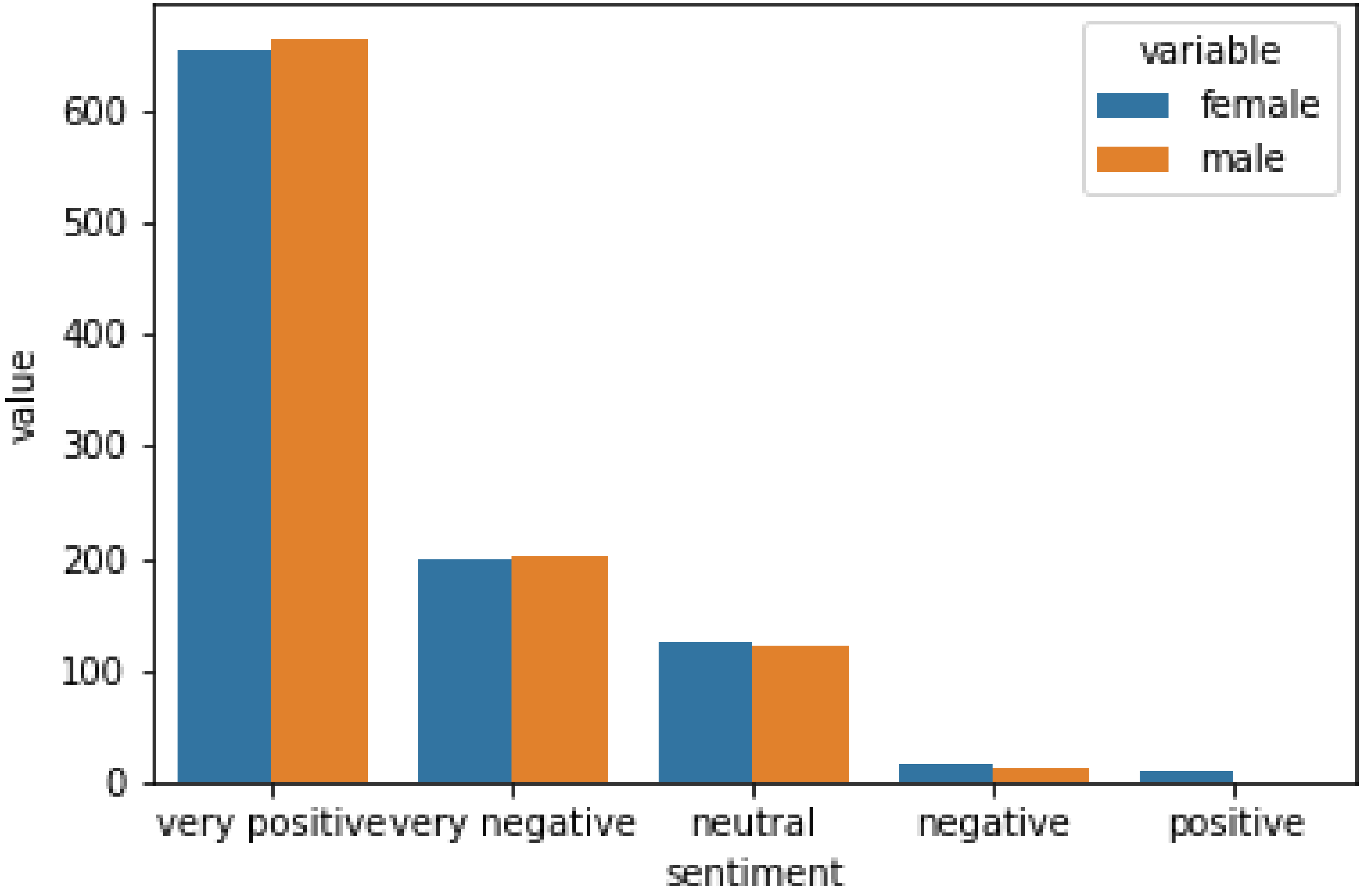
Results: Word Cloud (Male Subjects)

- Consider Top-5 predicted words for the [MASK] word
- The word cloud of predicted words for male Subjects



Results: Sentiment Analysis

- Sentiments associated with Female vs Male subject sentences



Conclusion & Future Work

- Analyzed Machine Translation (Persian to English) & Persian Language Models
- Our analysis clearly showed that gender bias occurs in translations from Persian to English
- Analysis of LM showed that there is no significant difference between sentiments of male vs female subject sentences.
- May conclude that in the translation, most of the bias is coming from the decoder language model which is trained on English corpus.
- The bias could be associated with the data.
- Simple Fix such as Rewriting. But not always useful. Requires Further study to reduce bias in LM.

References

[1] Assessing gender bias in machine translation: a case study with Google Translate by Pedro Henrique da Costa Avelar & Luís C. Lamb (https://www.researchgate.net/publication/332030363)

[2]Gender Bias in BERT - Measuring and Analysing Biases through Sentiment Rating in a Realistic Downstream Classification Task by Sophie F. Jentzsch et.al.

[3] Huggingface Library: https://huggingface.co/

JOHNS HOPKINS

WHITING SCHOOL of ENGINEERING