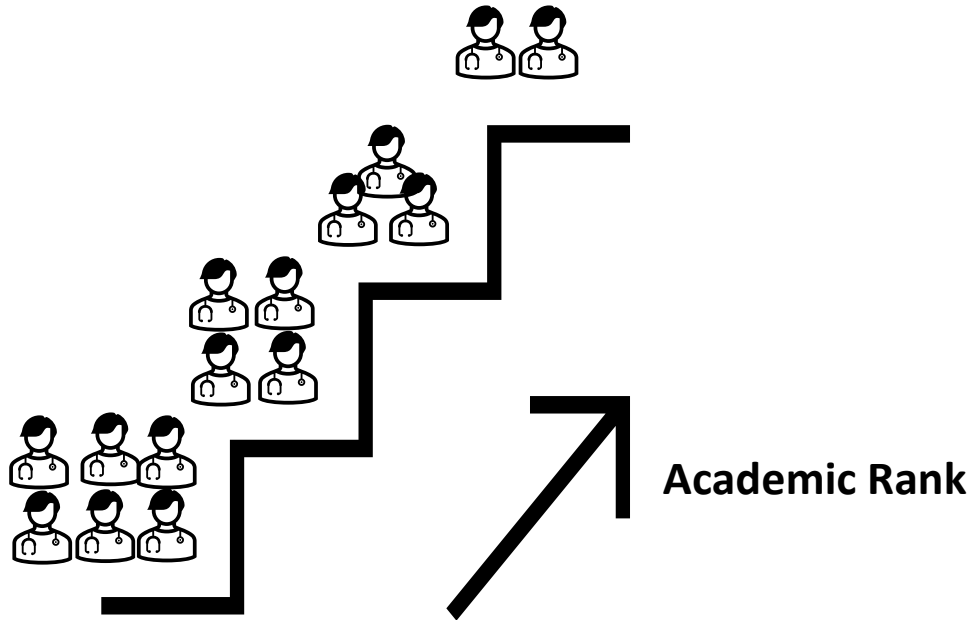


# Investigation of Implicit Gender Bias in Radiology Fellowship Letters of Recommendation Using Large Language Models

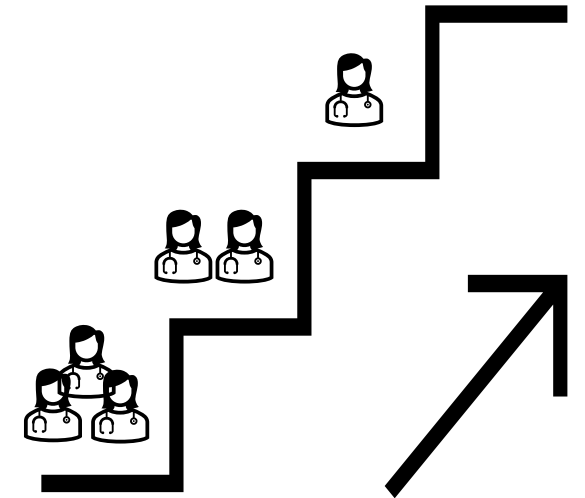
Alperen Elek, Sude Yildiz, Nisa Cem Oren, Lindsay Yang, Carina Yang, Benan Akca, **Batuhan Gundogdu**



# BACKGROUND

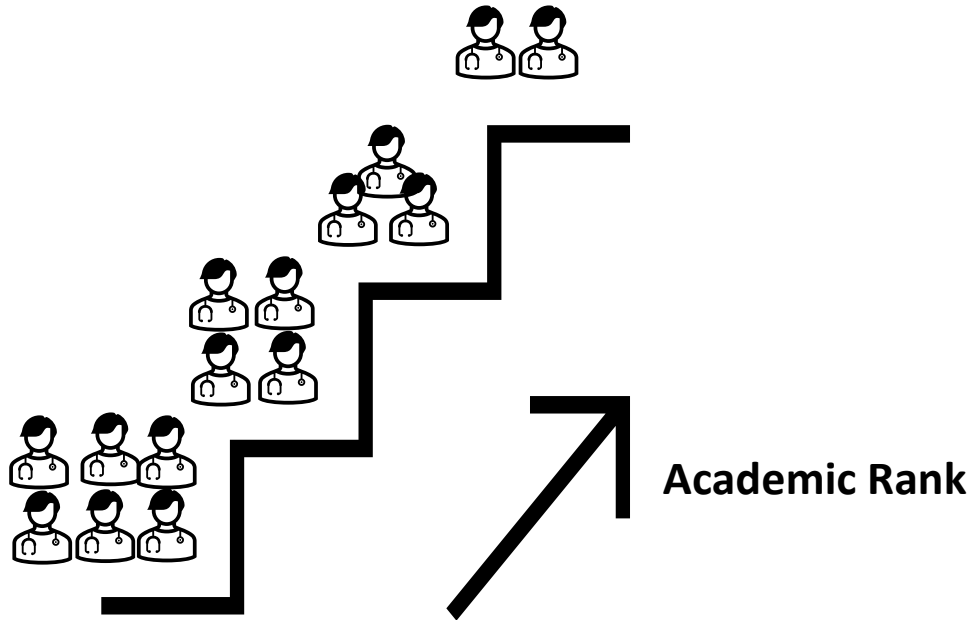


78% male



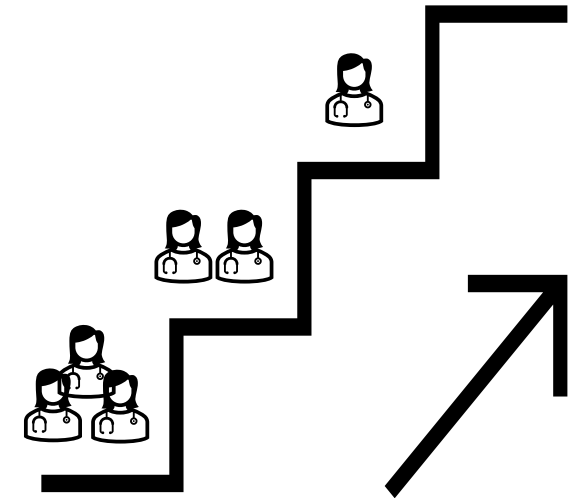
22% female

# BACKGROUND



**78% male**

*Gender bias in radiology ?*



**22% female**

# PURPOSE

The assess for *implicit* biases in letters of recommendations.

How are we going to *measure* ?

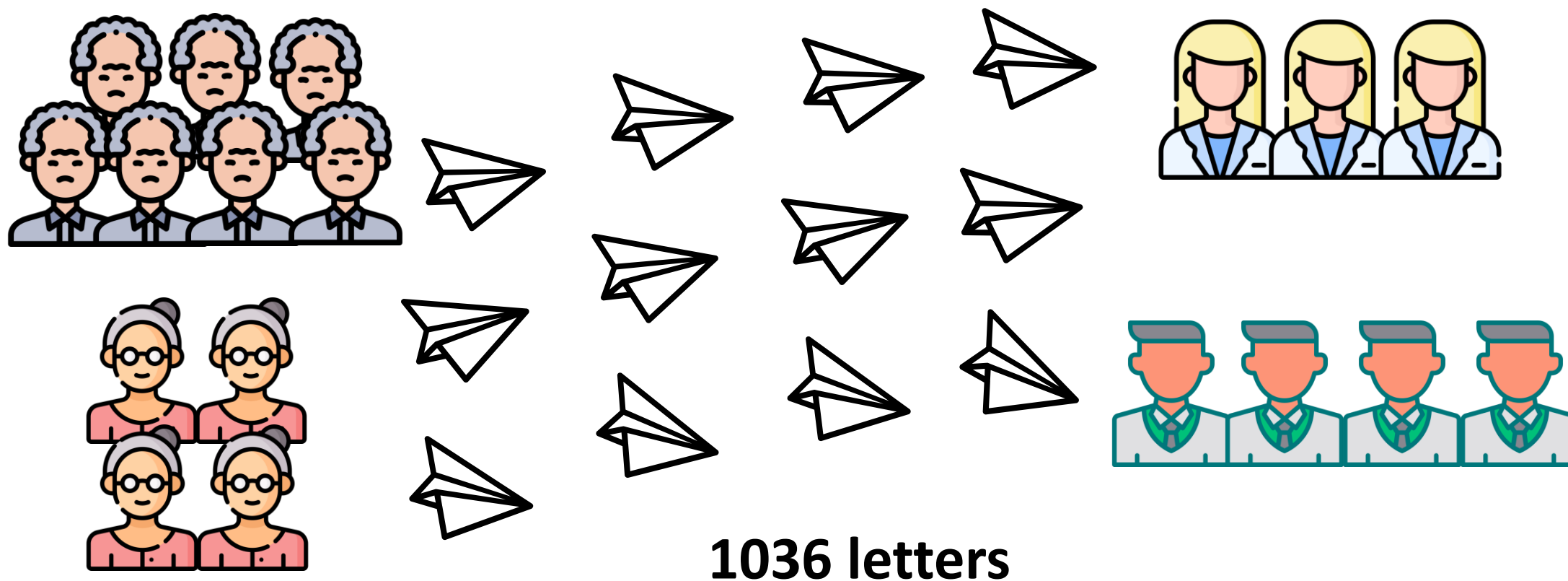
# PURPOSE

The assess for *implicit* biases in letters of recommendations.

How are we going to *measure* ?

**Perplexity Score of the LLM**

## Letters of Recommendation for Fellowship at University of Chicago



- Texts were anonymized and converted from PDF to text format for analysis.

# Semantics Comparison



**Semantic positivity score**

# Semantics Comparison



**Semantic positivity score**



**Agentic vs communal language**



# Semantics Comparison



**Semantic positivity score**



**Agentic vs communal language**

**No statistical difference between Male and Female candidates**

# Semantics Comparison



**Semantic positivity score**



**Agentic vs communal language**

**No statistical difference between Male and Female candidates**

**BUT...**

# Personalness of Letters



"On several occasions, Dr. John Doe's contributions allowed us to avoid case-delaying chart reviews or phone calls, making the day run more effectively."

A unique and personalized sentence

"I am confident that she has the skills to become a superb radiologist and I wish her the best of luck in her career."

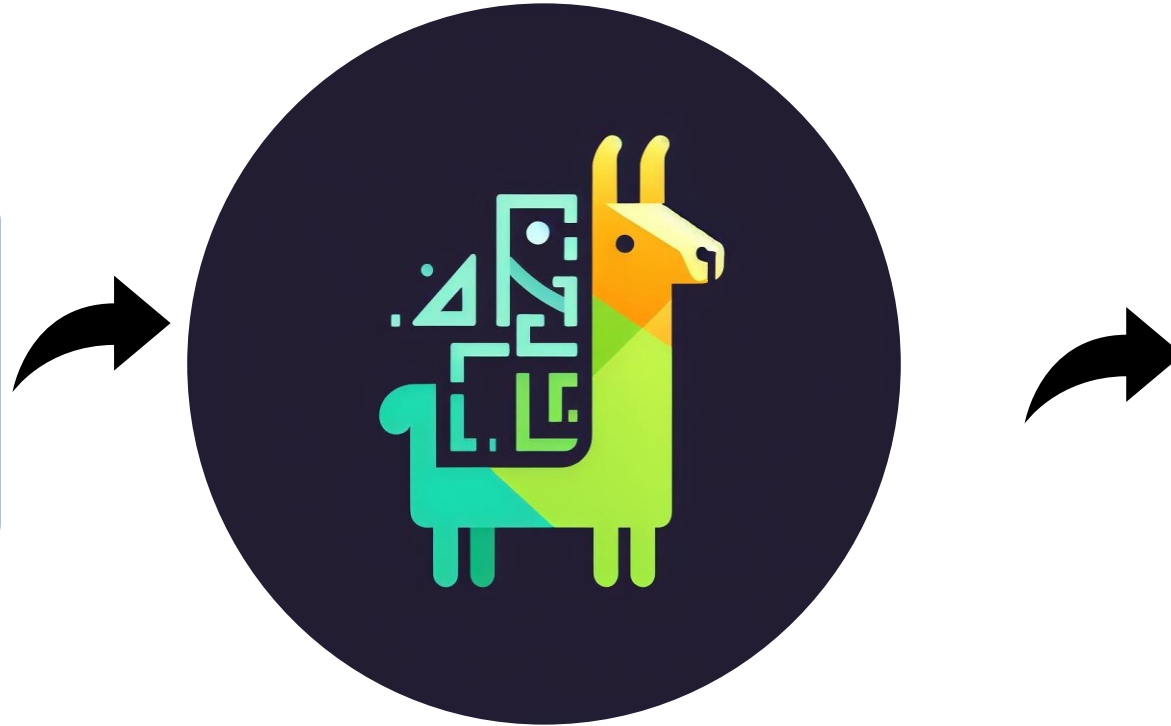


A generic, noninformative sentence that we always see in letters of recommendations

# Letter Uniqueness Calculation

**PPL Score**

“On several occasions, Dr. John Doe’s contributions allowed us to avoid case-delaying chart reviews or phone calls, making the day run more effectively.”

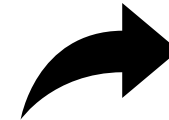
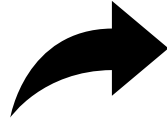


**80.97**

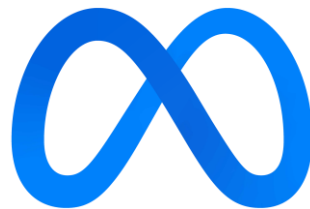
# Letter Uniqueness Calculation

**PPL Score**

"I am confident that she has the skills to become a superb radiologist and I wish her the best of luck in her career."

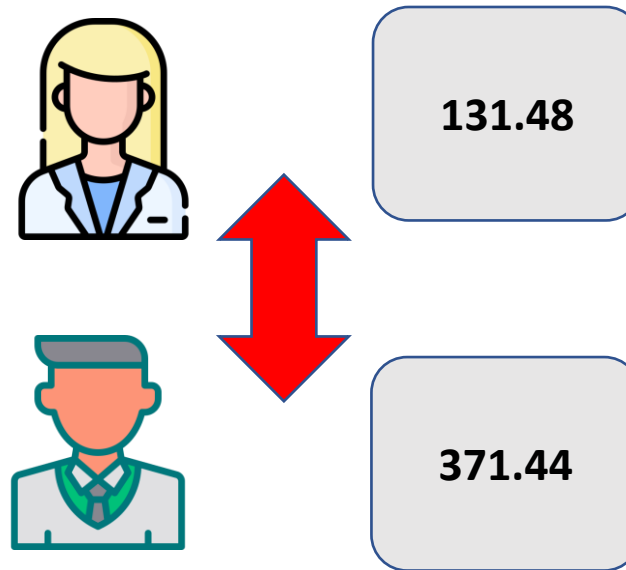


**4.52**



# RESULTS

PPL scores of letters written for male applicants were significantly higher than letters written for female applicants ( $p < 0.05$ ), indicating **more personalized** letters **for male applicants** and **more generic** letters **for female applicant**.



## Conclusions

- Recent research indicates a decrease in gender bias in academia, but potential **implicit bias** persists in reference letters.
- Analyzing perplexity values through **large language models can help** detect and address bias in human-written texts.

# Examples

High PPL Score (Potentially more personal, experience-oriented and focused on the candidate)	Low PPL Score (Potentially more generic, not conveying personal experience – despite being powerful and very positive)
“I have seen him delay his drive home from work because he helped a busy transporter bywheeling a patient back upstairs.”	“Please feel free to contact me if you have any questions regarding this letter.”
“On several occasions, Dr. John Doe’s* contributions allowed us to avoid case-delaying chart reviews or phone calls, making the day run more effectively.”	“I am confident that she has the skills to become a superb radiologist and I wish her the best of luck in her career.”
“John Doe is a leader and has thrived in situations requiring leadership throughout his academic career.”	“She is a kind person who goes out of his way to help others around him.”
“From the outset of his residency, John Doe stood out amongst his peers as a mature and leadership-oriented resident, with a unique aptitude for radiology.”	“He is very intelligent, has a good eye for radiologic findings, and has a natural ability to communicate complicated radiologic findings to consulting physicians and patients.”
“More remarkably, he fastidiously created electronic flashcards notes on every major topic in the text.”	“Should you have any questions or would like to further discuss his candidacy for your program, please feel free to contact me.”
“His presentations was outstanding - comprehensive in scope yet presented efficiently and effectively.”	“She has been one of the easiest people to work with, always showing up on time and ready for the day.”
“He always asked insightful questions after each procedure.”	“She has a great sense of humor and is very enjoyable to work with on a daily basis.”
“From the outset of his residency, Dr. John Doe stood out amongst his peers as a mature and leadership-oriented resident, with a unique aptitude for radiology”	
“John Doe is an extremely motivated individual and has clearly taken ownership of his education during residency.”	

The applicant names are anonymized to John Doe and Jane Doe for privacy.

The sentences provided in this Table are examples taken from reference letters. The sentences with low perplexity scores fall within the range of 5.3 to 11.9, whereas the sentences with high perplexity scores fall within the range of 100.7 to 327.4



Thank you!



## References

- Touvron, Hugo, et al. "Llama 2: Open foundation and fine-tuned chat models." *arXiv preprint arXiv:2307.09288* (2023).
- Colla D, Delsanto M, Agosto M, Vitiello B, Radicioni DP. Semantic coherence markers: The contribution of perplexity metrics. *Artif Intell Med.* 2022 Dec;134:102393. doi: 10.1016/j.artmed.2022.102393. Epub 2022 Sep 5. PMID: 36462890.
- Kubik-Huch RA, Vilgrain V, Krestin GP, Reiser MF, Attenberger UI, Muellner AU, Hess CP, Hricak H. Women in radiology: gender diversity is not a metric-it is a tool for excellence. *Eur Radiol.* 2020 Mar;30(3):1644-1652. doi: 10.1007/s00330-019-06493-1. Epub 2019 Dec 4. PMID: 31802213; PMCID: PMC7033068.