








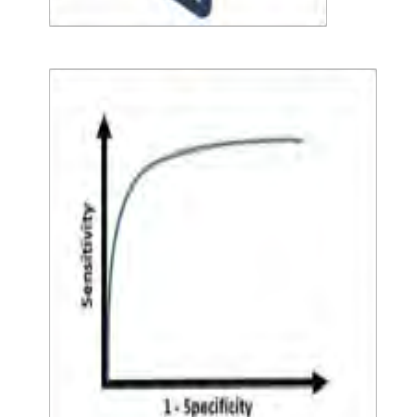
# Advancing Holistic Review During Residency Application: Using Natural Language Processing to Applicants' Experiences To Predict an Interview Invitation

Arun Umesh Mahtani, MD, MS<sup>1,2</sup>, Ilan Reinstein, MS<sup>1</sup>, Marina Marin, MS<sup>1</sup>, Jesse Burk-Rafel, MD, MRes<sup>1</sup>  
Email: [arun.mahtani@nyu.edu](mailto:arun.mahtani@nyu.edu) | <sup>1</sup> NYU Grossman School of Medicine, <sup>2</sup> Richmond University Medical Center/Mount Sinai

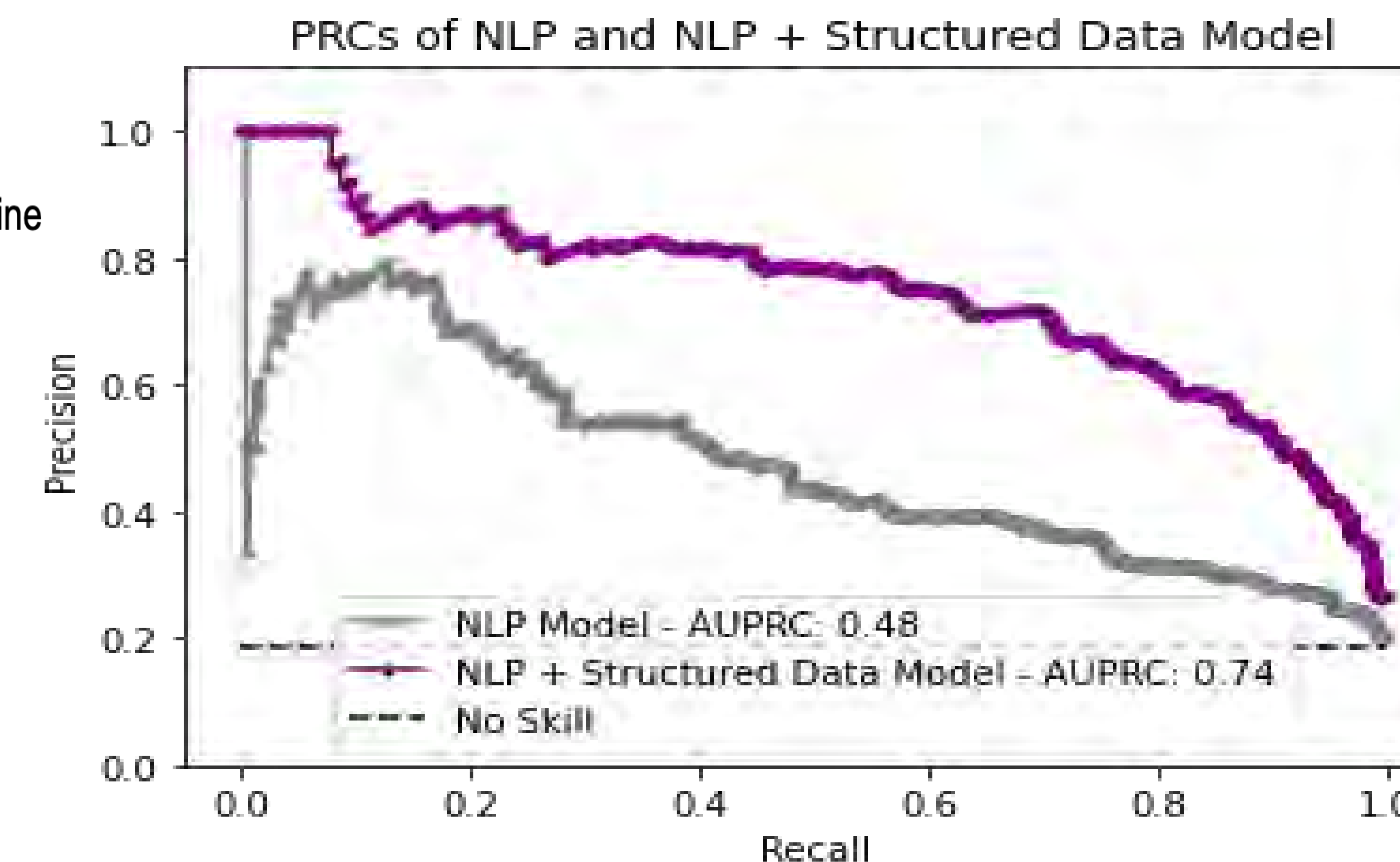
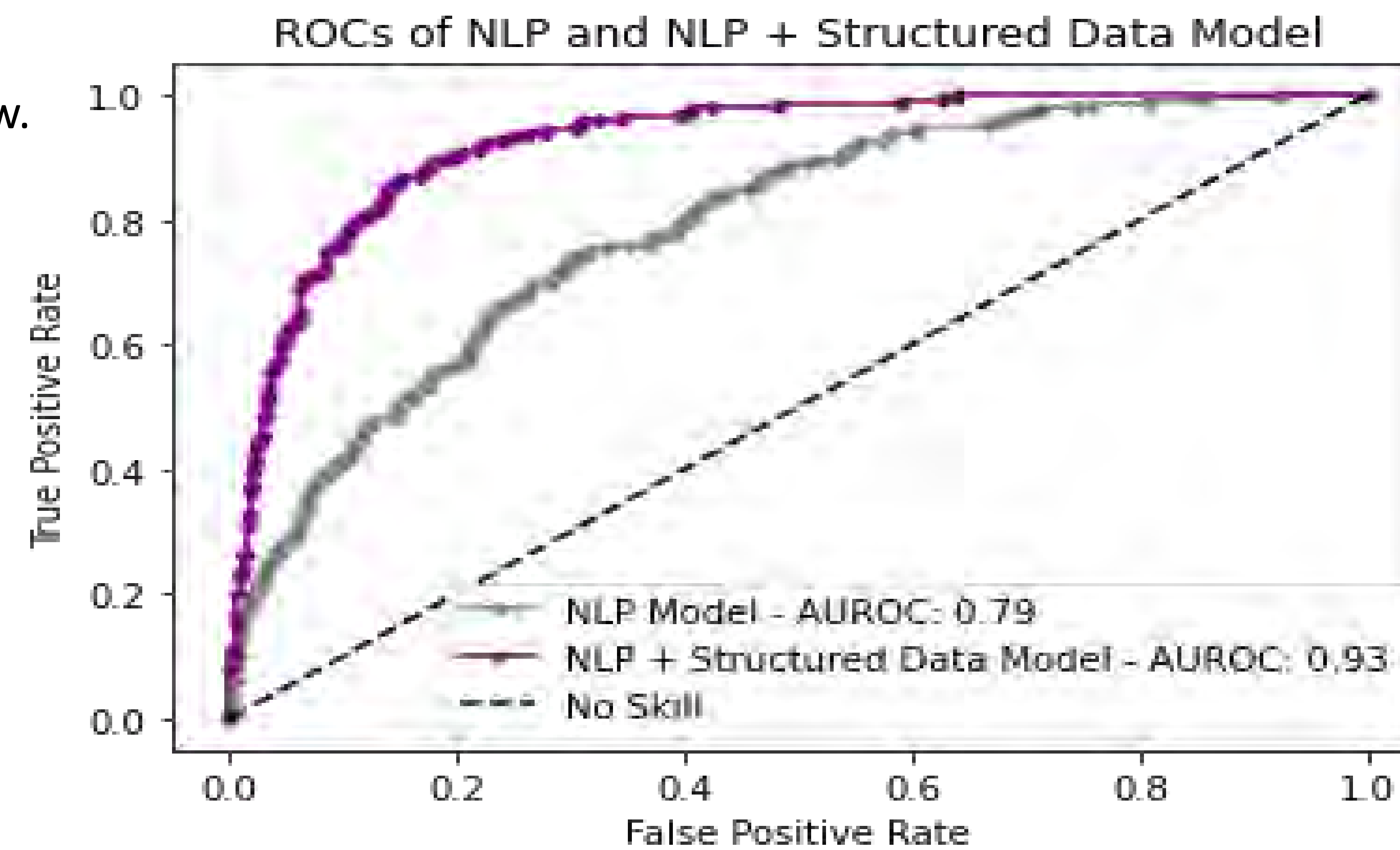
## INTRODUCTION

- Residency programs face increasing application numbers, limiting their ability to conduct holistic review.
- Efforts to increase holistic review and help programs appraise applicants are ongoing.<sup>1,2</sup>
- Assessing narrative components of residency applications remains a bottleneck.
- Programs resort to filter-based academic metrics and attributes that are known to have racial, gender, and other inequities.<sup>3-5</sup>

## METHODS

-  6500 ERAS applications from 2017-2019
-  Extracted experiences from each AAMC ID
-  Experiences preprocessed using spaCy's en\_core\_web\_lg pipeline
-  Word importance via TFIDF
-  Filtered words were split into train and test set
-  Logistic regression model using NLP & NLP + structured data
-  Features selected with nonzero coefficients. Each applicant was assigned a predicted probability score
-  Model performance assessed

## RESULTS



## DISCUSSION

- Higher odds of an interview invitation linked with:
  - ✓ Terms signifying active research, active leadership roles, or work associated with underserved communities and health disparities.
  - ✓ Experiences from highly ranked medical school in NE region.
- Heavy reliance on structured data for initial screening of applications.
- Most important predictor of combined model based on NLP of the experiences
- Promise of an NLP-based approach based on:
  - ✓ Moderately strong predictive value of the NLP-only model
  - ✓ Persistence of the NLP predictor in the combined model

## NEXT STEPS

- Apply advanced NLP approaches and ML classifiers to analyze all narrative residency application components:
  - ✓ Personal statements
  - ✓ Medical Student Performance Evaluations
  - ✓ Letters of recommendation.
- Engage other programs that utilize the experience entries more heavily in applicant screening to gather validity evidence for an NLP-based approach.
- Develop open-source universal pipeline for any program to analyze narrative components of applications.

## REFERENCES

- 1) Vela MB, Chin MH, Peek ME. Keeping Our Promise - Supporting Trainees from Groups That Are Underrepresented in Medicine. *N Engl J Med.* Aug 5 2021;385(6):487-489.
- 2) Supplemental ERAS application (for the ERAS 2022 cycle). AAMC Accessed 8/19, 2021. <https://students-residents.aamc.org/applying-residencies-eras/supplementalerasapplication>
- 3) Capers Qt, Clinchot D, McDougale L, Greenwald AG. Implicit Racial Bias in Medical School Admissions. *Acad Med.* Mar 2017;92(3):365-369.
- 4) Maxfield CM, Thorpe MP, Desser TS, et al. Bias in Radiology Resident Selection: Do We Discriminate Against the Obese and Unattractive? *Acad Med.* Nov 2019;94(11):1774-1780.
- 5) Filippou P, Mahajan S, Deal A, et al. The Presence of Gender Bias in Letters of Recommendations Written for Urology Residency Applicants. *Urology.* Dec 2019;134:56-61.