

# Can Artificial Intelligence Fool Residency Selection Committees? Analysis of Personal Statements by Real Applicants and Generative AI, A Randomized, Single-Blind Multi-Center Study

## Authors

Zachary C. Lum DO<sup>1</sup>  
Lohitha Guntupalli BS<sup>1</sup>  
Augustine M. Saiz MD<sup>2</sup>  
Holly Leshikar MD<sup>3</sup>  
Hai V. Le MD<sup>2</sup>  
John P. Meehan MD<sup>2</sup>  
Eric G. Huish DO<sup>3</sup>

## Institutions

<sup>1</sup>Nova Southeastern University, Kiran Patel School of Osteopathic and Allopathic Medicine, Department of Surgery, Davie, Florida.

<sup>2</sup>University of California: Davis, School of Medicine, Department of Orthopaedic Surgery, Sacramento, California.

<sup>3</sup>San Joaquin General Hospital, French Camp, California.

## Author Disclosures

Zachary C. Lum DO<sup>1</sup>  
American Association of Hip and Knee Surgeons: Board or committee member  
JAOAO: Editorial or governing board  
Lohitha Guntupalli BS<sup>1</sup> – None  
Augustine M. Saiz MD  
Orthopaedic Research Society: Board or committee member  
Orthopaedic Trauma Association: Board or committee member  
Holly Leshikar MD – None  
Hai V. Le MD – None  
John P. Meehan MD  
DePuy, A Johnson & Johnson Company: Paid presenter or speaker; Research support  
OrthoAlign: Paid consultant  
Eric G. Huish DO  
Synthes: Paid Consultant

## Introduction:

The potential capabilities of generative artificial intelligence (AI) tools have been relatively unexplored, particularly in the realm of creating personalized statements for medical students applying to residencies. This study aimed to investigate the ability of generative AI, specifically ChatGPT and Google BARD, to generate personal statements and assess whether faculty on residency selection committees could evaluate differences between real and AI statements.

## Methods:

15 REAL personal statements were used to generate 15 unique and distinct personal statements from ChatGPT and BARD each, resulting in a total of 45 statements. Statements were then randomized, blinded, and presented to a group of faculty reviewers on residency selection committees. Reviewers assessed the statements by 14 metrics including if the personal statement was AI-generated or real. Comparison of all metrics was performed.

## Results:

Faculty correctly identified 88% (79/90) real statements, 90% (81/90) BARD and 44% (40/90) ChatGPT statements. Accuracy of identifying REAL and BARD statements was 89%, but this dropped to 74% when including ChatGPT. Additionally, the accuracy did not increase as faculty members reviewed more personal statements (AUC 0.498, p=0.966).

BARD performed poorer than both REAL and ChatGPT across all metrics (p<0.001). Comparing real with ChatGPT, there were no difference in most metrics, except for Personal Interests, Reasons for Choosing Residency, Career Goals, Compelling Nature and Originality, and all favoring the real personal statements (p=0.05, p=0.028, p=0.015, p=0.001, p=0.001, respectively).

## Discussion:

Faculty members accurately identified real and BARD statements, but ChatGPT deceived them 56% of the time. Although AI can craft convincing statements that are sometimes indistinguishable from real ones, replicating the humanistic experience, personal nuances and individualistic elements found in real personal statements is difficult. Residency selection committees might want to prioritize these particular metrics while assessing personal statements, given the growing capabilities of AI in this arena.

## Significance/Clinical Relevance:

With the rise of generative artificial intelligence tools, the potential for its use in crafting personal statements for medical residency has not been previously studied. ChatGPT was able to craft convincing personal statements which were at times indistinguishable from real ones, though individualistic elements were difficult to replicate.