Research Study

Who Writes AI Impressions Best?

Human Versus AI

Rad AI

The Problem

Radiologists read nearly 100,000 words a week. These long, detailed reports are cognitively demanding, contributing to a 44-65% burnout rate[1], leading to early retirement and workforce shortages.

The Study

Conducted in collaboration with researchers at Moffitt Cancer Center, the analysis included 200 de-identified CT oncologic reports dictated by four abdominal radiologists — 50 per radiologist. There were three impression types: human-authored, custom AI-generated (Rad AI Impressions) and generic AI-generated.

The Findings

Word Count

Humans: ~ 41 words
Custom AI: ~ 34
Generic AI: ~ 75

Generic AI tended to be more wordy, while humans and custom AI kept it shorter and more to the point.

Impression Items*

Humans: ~ 3
Custom AI: ~ 3
Generic AI: ~ 6

Generic AI often listed more impression items than necessary, while humans and custom AI focused on the key points.

*impression items = bullet points or numbered items in a list

Mean Generation Time

Custom AI: ~ 2 seconds
Generic AI: ~ 12 seconds

Custom AI was much quicker, producing results in just a couple of seconds, while generic AI took noticeably longer.

Radiologist and Oncologist Preferences

Radiologists preferred human and custom AI over generic AI ($p < 0.001$). Oncologists were more evenly split, with a slight preference for custom AI ($p=0.21$-0.63).

Takeaway

The winner isn't human or AI; it's custom AI plus humans working together.

"Impressions are the most critical part of the report.

And we found that in addition to accuracy, customization mattered to the radiologist when it comes to them feeling confident that the AI clearly and effectively communicates the information in a way that will align with clinical workflows."

-**Andrew Del Gaizo, MD,**
one of the study authors and CMIO at Rad AI

Citation

1. American College of Radiology. (n.d.). Burnout fueling workforce woes. https://www.acr.org/Clinical-Resources/Publications-and-Research/ACR-Bulletin/Burnout-Fueling-Workforce-Woes
2. Phadke, S., Suresh, N., Allen, Z., Balagopal, A., Chan, S., Shah, A., Winter, M., Lam, C., Rose, T., Araujo, C., Ahmed, A., Imanirad, I., Berland, L., & Del Gaizo, Al. (2026). Comparison of AI generated radiology impressions: a multi-stakeholder evaluation. npj Digital Medicine. https://www.nature.com/articles/s41746-026-02586-6