

54. **MAN VERSUS MACHINE: CAN ARTIFICIAL INTELLIGENCE DISTINGUISH BETWEEN THE QUALITY OF ONLINE HEALTH INFORMATION ON SOCIAL MEDIA?**

Christian A. Persichetti<sup>1</sup>.

<sup>1</sup> Fourth Year Medical Student. Rutgers New Jersey Medical School, Newark, United States of America.

**BACKGROUND:** ChatGPT is a popular artificial intelligence service offered by OpenAI. It's utility in assessing quality of online medical information has been investigated by Golan and Ripps et. al. In July 2023, ultimately concluding that ChatGPT was unable to discern poor from high quality information. However, this study focused on information from websites generated from a Google search. To the author's knowledge, there has been no evaluation of the ability of ChatGPT to stratify the quality of online medical information from social media videos. Two previous studies, by Dhami and Wescott et. al. In May 2023 and Rehman et. al. In 2021, identified differences via manual calculation of DISCERN scores between physician and non-physician Tik-Tok videos on the topic of hidradenitis suppurativa. This study aims to identify whether ChatGPT can distinguish between physician and non-physician health videos pertaining to hidradenitis suppurativa on Tik-Tok via DISCERN score. **METHODS:** A similar methodology to Dhami and Wescott et. al. was used in an attempt to capture the same population of videos. A new Tik-Tok account was created to eliminate any previous search bias. The term "#hidradenitissuppurativa" was searched in July, 2023. Videos were filtered by "most-liked." Inclusion criteria consisted of videos posted prior to October 25, 2022, that did not satisfy the exclusion criteria. Exclusion criteria consisted of videos in a language other than English, duplicated videos, videos that did not offer health advice, or videos deemed irrelevant to hidradenitis suppurativa by the author. Videos included were transcribed by the website TokScript (<https://script.tokaudit.io>) or by the author if transcripts were not available by that method. ChatGPT was prompted to evaluate transcripts based on the DISCERN criteria, which were copied directly from the Discern Instrument website "Rating this question" pop-up. Each of the 16 question criteria were submitted to ChatGPT separately. Transcriptions were provided to ChatGPT and a DISCERN score was generated. A Shapiro-Wilk test was performed on the sample of DISCERN scores to determine normality. A Mann-Whitney U test was performed between physician and non-physician DISCERN scores. **RESULTS:** Of 377 Tik Tok videos returned with the search, 43 satisfied inclusion criteria (11.4%). Of these, 19 (44.2%) were produced by a physician and 24 (55.8%) by a non-physician. Aggregate DISCERN scores were not normally distributed (Shapiro-Wilk  $p < 0.0002$ ). Physician produced videos had a mean DISCERN score of 41.87 with an interquartile range of 26-57. Mean DISCERN score for non-physician produced videos was 24.56 with an interquartile range of 16-30.5. Mann-Whitney U test returned a  $p < 0.0006$ . **CONCLUSION:** This study suggests that ChatGPT may have the capacity to distinguish high from low quality social media videos to some extent based on the DISCERN score. More evaluation is necessary to elucidate the boundaries of utilizing ChatGPT in this fashion. This has the potential to suggest that polishing this utilization may automate this process in the future, saving valuable time for human researchers in this field.

**Key words:** Ai Artificial Intelligence; Consumer Health Information; Social Media (Source: MeSH-NLM).