

NOTEWORTHY, SOFTWARE

FIGHTING AI WITH AI

27.04.2025

Here is our newest paper – a nice collaboration with [Andrea Taloni et al.](#) along [with a nice commentary](#) – to recognize [surgisphere-like](#) fraud

Recently, it was proved that the large language model Generative Pre-trained Transformer 4 (GPT-4; OpenAI) can fabricate synthetic medical datasets designed to support false scientific evidence. To uncover statistical patterns that may suggest fabrication in datasets produced by large language models and to improve these synthetic datasets by attempting to remove detectable marks of nonauthenticity, investigating the limits of generative artificial intelligence.

[...] synthetic datasets were produced for 3 fictional clinical studies designed to compare the outcomes of 2 alternative treatments for specific ocular diseases. Synthetic datasets were produced using the default GPT-4o model and a custom GPT. Data fabrication was conducted in November 2024. Prompts were submitted to GPT-4o to produce 12 “unrefined” datasets, which underwent forensic examination. Based on the outcomes of this analysis, the custom GPT Synthetic Data Creator was built with detailed instructions to generate 12 “refined” datasets designed to evade authenticity checks. Then, forensic analysis was repeated on these enhanced datasets. [...]

Sufficiently sophisticated custom GPTs can perform complex statistical tasks and may be abused to fabricate synthetic datasets that can pass forensic analysis as authentic.

