

**REVIEW JURNAL
PENELITIAN EKSPERIMENTAL
MATA KULIAH RISET INFORMATIKA**

“Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews”



Andini Fitriyah Salsabilah (21081010217)

DOSEN PENGAMPU:
Assoc. Prof. Dr. Basuki Rahmat, S.Si, MT, ITS-AI

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA
TIMUR 2024**

Penelitian ini termasuk penelitian eksperimental karena penelitian ini menguji pendekatan baru untuk mendeteksi konten yang dimodifikasi oleh AI pada skala besar dalam sistem peninjauan ilmiah. Peneliti menggunakan metode eksperimental dengan menghasilkan data pelatihan melalui simulasi teks yang ditulis manusia dan yang dihasilkan AI. Mereka kemudian menerapkan *maximum likelihood estimation* (MLE) untuk mengestimasi proporsi teks yang kemungkinan besar dihasilkan oleh model AI, seperti ChatGPT, dalam corpus ulasan ilmiah. Eksperimen ini mengkaji efek penggunaan ChatGPT dalam ulasan konferensi, termasuk variasi pada waktu pengajuan dan korelasi dengan kepercayaan diri reviewer.

Judul	Predictive Model for Extended-Spectrum β -Lactamase-Producing Bacterial Infections Using Natural Language Processing Technique and Open Data in Intensive Care Unit Environment: Retrospective Observational Study
Jurnal	arXiv
Volume	-
Tahun	2024
Penulis	Weixin Liang, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, Haotian Ye, Sheng Liu, Zhi Huang, Daniel A. McFarland, James Y. Zou
Instansi Penulis	Stanford University, NEC Labs America, dan UC Santa Barbara.
Abstrak	Penelitian ini mengembangkan metode untuk mengestimasi proporsi teks yang dimodifikasi oleh AI dalam ulasan ilmiah, terutama setelah rilis ChatGPT. Metode ini menggunakan data pelatihan dari teks yang diketahui asalnya, baik manusia maupun AI, dan menerapkan <i>maximum likelihood estimation</i> (MLE) untuk mendeteksi teks yang dihasilkan AI pada tingkat populasi.
Permasalahan	Adanya peningkatan penggunaan AI seperti ChatGPT yang mengaburkan batas antara konten asli buatan manusia dan konten yang dihasilkan oleh AI, serta implikasinya terhadap proses ulasan ilmiah.
Subjek Penelitian	Ulasan ilmiah dari konferensi seperti ICLR, NeurIPS, CoRL, dan EMNLP, serta jurnal dari Nature Portfolio.
Metode Penelitian	Penelitian menggunakan metode eksperimental dengan MLE untuk mendeteksi proporsi teks yang dimodifikasi oleh AI pada skala besar, termasuk verifikasi melalui data semi-sintetis dan perbandingan dengan metode deteksi berbasis instance.
Hasil Penelitian	Studi menemukan bahwa antara 6,5% hingga 16,9% ulasan ilmiah setelah rilis ChatGPT berpotensi dimodifikasi secara substansial oleh AI, terutama mendekati batas waktu pengumpulan ulasan.
Kekuatan Penelitian	Memperkenalkan metode yang efisien untuk mendeteksi

	modifikasi AI pada tingkat populasi. Metode menunjukkan ketahanan terhadap variasi topik dan mampu memberikan estimasi yang akurat pada berbagai dataset.
Kelemahan Penelitian	Bergantung pada data pelatihan yang diketahui asalnya, yang mungkin tidak mewakili seluruh penggunaan AI yang ada. Kemungkinan adanya perbedaan dalam hasil karena variasi teknik prompt dan ketergantungan pada model AI tertentu (ChatGPT).
Kesimpulan	Metode ini menawarkan pendekatan yang menjanjikan untuk memantau konten yang dimodifikasi AI dalam ulasan ilmiah, dengan potensi aplikasi yang luas dalam ekosistem informasi lainnya. Hasil studi ini menyoroti kebutuhan untuk penelitian lebih lanjut dan penerapan metode ini di berbagai bidang untuk pemantauan konten berbasis AI.