

Susceptibility to Deception: Exploring the Cognitive Process of Perceiving and Discerning Deepfake AI Using Verbal Protocol Analysis

Adira Ridhiani Putri¹, Elga Andriana²

^{1,2}Faculty of Psychology, Universitas Gadjah Mada, Indonesia

e-mail: *¹adira.ridhiani2103@mail.ugm.ac.id, ²elga.andriana@ugm.ac.id

Abstract. Deepfake AI is a technology that allows for the synthesis of audiovisual media depicting people or scenarios that never existed. While machine-based strategies for detecting deepfake media are in continuous development, not many have explored the cognitive processes that individuals engage in when encountering deepfake content. This study seeks to understand the underlying cognitive mechanisms by which people perceive manipulated media and how susceptible they are to digital deception. Through utilising Verbal Protocol Analysis, this research aims to gain insights into the cognitive factors that influence the strategies individuals use to discern between deepfake AI and authentic media. This study explores deeper into the aspects of information processing, decision-making, and other influencing factors, such as socioemotional processing and media awareness, that play a role in human deepfake detection, highlighting the complex nature of discerning deepfakes and how advancements in this field could lead to challenges in our capability to detect deepfake AI.

Keywords: *deepfake AI, perception, cognitive process, deception, discerning strategies*

Abstrak. Deepfake AI adalah teknologi yang memungkinkan untuk sintesis media audiovisual untuk menggambarkan orang atau skenario yang tidak pernah ada. Meskipun strategi berbasis mesin untuk mendeteksi media deepfake terus dikembangkan, belum banyak yang mengeksplorasi proses kognitif yang dilalui individu saat disajikan konten deepfake. Penelitian ini bertujuan untuk memahami proses kognitif yang mendasari orang dalam menanggapi media yang dimanipulasi dan seberapa rentan mereka terhadap penipuan digital. Dengan menggunakan Analisis Protokol Verbal sebagai desain, penelitian ini bertujuan untuk mendapatkan pemahaman mengenai faktor-faktor kognitif yang memengaruhi strategi individu dalam membedakan antara deepfake AI dan media otentik. Penelitian ini memberikan pemahaman mengenai berbagai aspek pemrosesan informasi, pengambilan keputusan, dan faktor-faktor lain yang dapat memengaruhi, seperti pemrosesan sosioemosional dan kesadaran media yang berperan dalam deteksi deepfake pada manusia yang menyoroti sifat kompleks dalam membedakan deepfake dan bagaimana kemajuan dalam bidang ini dapat menimbulkan tantangan dalam kemampuan kita untuk mendeteksi deepfake AI.

Keywords: *deepfake AI, persepsi, proses kognitif, desepsi, strategi membedakan*