

**TRANPARANSI PENGGUNAAN DATA PRIBADI DALAM
PENGUMPULAN DATASET PADA PENGEMBANGAN *GENERATIVE
ARTIFICIAL INTELLIGENCE* BERDASARKAN UNDANG-UNDANG
NOMOR 27 TAHUN 2022 TENTANG PELINDUNGAN DATA PRIBADI**

Oleh
Elvira Clarista Faiqah
NIM. 2205040134

ABSTRAK

Perkembangan *Generative Artificial Intelligence* (GenAI) mendorong meningkatnya praktik pengumpulan data pribadi melalui teknik *web scraping* dan *web crawling* dalam pembentukan *dataset* AI. Praktik tersebut menimbulkan persoalan hukum, khususnya terkait penerapan prinsip transparansi terhadap subjek data pribadi sebagaimana diatur dalam Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi (UU PDP). Penelitian ini bertujuan untuk menganalisis penerapan prinsip transparansi dalam praktik pengumpulan *dataset* GenAI berdasarkan UU PDP serta akibat hukum atas pelanggaran transparansi dalam pengumpulan data pribadi untuk pengembangan GenAI. Penelitian ini menggunakan metode yuridis normatif dengan pendekatan perundang-undangan, dan pendekatan komparatif melalui perbandingan antara UU PDP dan *General Data Protection Regulation* (GDPR). Hasil penelitian ini menjelaskan praktik pengumpulan dataset untuk pengembangan GenAI yang dilakukan secara otomatis melalui teknik seperti *web scraping* dan *web crawling* menunjukkan bahwa penerapan prinsip transparansi sebagaimana diatur dalam Pasal 16 ayat (2) huruf a UU PDP belum dapat terlaksana secara optimal. Karakteristik pengumpulan data yang bersifat otomatis, masif, dan sering kali dilakukan tanpa interaksi langsung dengan subjek data menyebabkan sulitnya pemberian informasi yang jelas mengenai pemrosesan data pribadi. Dengan demikian, meskipun UU PDP telah mewajibkan adanya transparansi dalam pemrosesan data pribadi, pengaturan tersebut belum sepenuhnya mampu mengakomodasi praktik pengumpulan data otomatis dalam pengembangan GenAI, sehingga pemenuhan hak transparansi subjek data masih menghadapi kendala dalam penerapannya. Selain itu, pelanggaran terhadap prinsip transparansi dapat menimbulkan akibat hukum bagi pengendali data apabila menyebabkan kerugian bagi subjek data pribadi.

Kata Kunci: *Generative Artificial Intelligence* (GenAI), Prinsip Transparansi, Pelindungan Data Pribadi.

**TRANSPARENCY OF PERSONAL DATA USAGE IN DATASET
COLLECTION IN THE DEVELOPMENT OF GENERATIVE ARTIFICIAL
INTELLIGENCE BASED ON LAW NUMBER 27 OF 2022 CONCERNING
PERSONAL DATA PROTECTION**

**By
Elvira Clarista Faiqah
NIM. 2205040134**

ABSTRACT

The development of Generative Artificial Intelligence (GenAI) has led to an increase in the collection of personal data through web scraping and web crawling techniques for the creation of AI datasets. These practices raise legal issues, particularly regarding the application of the principle of transparency toward data subjects, as stipulated in Law No. 27 of 2022 on Personal Data Protection (PDP Law). This study aims to analyze the application of the principle of transparency in GenAI dataset collection practices based on the PDP Law, as well as the legal consequences of transparency violations in the collection of personal data for GenAI development. This study employs a normative legal methodology using a statutory approach, as well as a comparative approach by comparing the PDP Law with the General Data Protection Regulation (GDPR). The results of this study reveal that the practice of collecting datasets for GenAI development—which is carried out automatically through techniques such as web scraping and web crawling—indicates that the application of the principle of transparency as stipulated in Article 16(2)(a) of the PDP Act has not yet been optimally implemented. The nature of data collection—which is automated, massive in scale, and often conducted without direct interaction with data subjects—makes it difficult to provide clear information regarding the processing of personal data. Thus, although the PDP Act mandates transparency in the processing of personal data, these regulations have not yet been fully able to accommodate automated data collection practices in GenAI development, meaning that the fulfillment of data subjects' right to transparency still faces obstacles in its implementation. Furthermore, violations of the principle of transparency may result in legal consequences for data controllers if they cause harm to data subjects.

Keywords: Generative Artificial Intelligence (GenAI), Transparency Principle, Personal Data Protection.