

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Kecerdasan dalam komputer merupakan kemampuan untuk meniru proses berpikir manusia, seperti belajar, menalar, memahami, dan mengambil keputusan dengan cara mengumpulkan, memproses, serta menganalisis informasi secara otomatis untuk mencapai tujuan tertentu secara efisien dan adaptif.¹ Salah satu cabang utama dari kecerdasan komputer adalah kecerdasan buatan atau yang biasa dikenal sebagai *Artificial Intelligence* yang selanjutnya akan disingkat menjadi AI. Menurut Russell dan Norvig kecerdasan buatan merupakan cabang dari ilmu komputer yang berfokus pada penciptaan sistem yang mampu berpikir dan bertindak secara rasional layaknya manusia.² AI memungkinkan komputer untuk melakukan proses kognitif, seperti belajar dari pengalaman, menalar berdasarkan data yang tersedia, dan mengambil keputusan yang optimal terhadap situasi tertentu. Dengan kata lain, AI adalah bentuk implementasi dari kecerdasan komputer yang meniru kemampuan manusia melalui algoritma dan model matematis yang dirancang untuk menyesuaikan diri dengan perubahan lingkungan dan tujuan yang diinginkan.³

Pada praktiknya, kecerdasan buatan dapat diwujudkan dalam bentuk perangkat lunak yang berdiri sendiri maupun terintegrasi ke dalam perangkat keras

¹ Joseph Teguh Santoso, *Kecerdasan Buatan (Artificial Intelligence)* (Semarang: Yayasan Prima Agus Teknik, 2023), hlm. 2.

² Stuart J. Russel dan Peter Norvig, *Artificial Intelligence: A Modern Approach* (Pearson Education, 2021).

³ *Ibid.*, hlm. 5.

atau mesin yang dirancang untuk menjalankan fungsi tertentu.⁴ Saat ini, AI telah mengalami perkembangan yang pesat dan mulai diterapkan di berbagai sektor. Namun, sebelum mencapai tahap implementasi tersebut, *Artificial Intelligence* telah melalui sejarah panjang dan proses pengembangan yang kompleks. Istilah kecerdasan buatan atau *Artificial Intelligence* mulai dikenal pada dekade 1950-an, tepatnya melalui forum ilmiah *Dartmouth Summer Research Project on Artificial Intelligence* yang diselenggarakan pada tahun 1956.⁵ Forum ini menjadi tonggak awal bagi pengembangan AI sebagai disiplin ilmu tersendiri, dengan fokus pada penciptaan mesin yang mampu meniru kecerdasan manusia melalui pemrosesan simbolik dan pembelajaran otomatis.⁶

Seiring meningkatnya kebutuhan terhadap sistem yang mampu meniru cara berpikir manusia, penelitian mengenai AI pun terus mengalami kemajuan. Berbagai pendekatan dan metode dikembangkan, mulai dari logika simbolik hingga pembelajaran mesin (*Machine Learning*), yang memungkinkan komputer untuk mengenali pola, membuat prediksi, dan mengambil keputusan secara mandiri. Perkembangan ini tidak hanya mencerminkan kemajuan dalam bidang ilmu komputer, tetapi juga menunjukkan pergeseran paradigma dalam memahami dan mereplika proses berpikir manusia melalui teknologi.⁷

Melalui pengembangan algoritma pembelajaran mesin (*machine learning*) dan *deep learning*, AI kini mampu mempelajari data dalam skala besar secara

⁴ Ari Juliano Gema, "Masalah Penggunaan Ciptaan Sebagai Data Masukan Dalam Pengembangan Artificial Intelligence Di Indonesia," *Technology And Economic Law Journal* 1, no. 1 (11 Februari 2022): 1–17, <https://doi.org/10.21143/TELJ.vol1.no1.1000>.

⁵ *Ibid.*, hlm. 2.

⁶ *Ibid.*, hlm. 2

⁷ *Ibid.*, hlm. 3.

otomatis. Salah satu inovasi terbesar dari kemajuan ini adalah munculnya *Generative Artificial Intelligence* atau yang selanjutnya disebut GenAI. GenAI merupakan salah satu bidang dalam kecerdasan buatan yang menitikberatkan pada pengembangan sistem yang mampu menciptakan data baru dengan mengacu pada pola-pola yang telah dipelajari sebelumnya.⁸ Jenis data yang dihasilkan dapat berupa teks, gambar, audio, video, maupun bentuk konten digital lainnya yang menyerupai data asli.

Berbeda dengan pendekatan tersebut, kecerdasan buatan tradisional atau biasa disebut sebagai *discriminative Artificial Intelligence* lebih diarahkan untuk menganalisis dan mengklasifikasikan data berdasarkan pola yang telah ada.⁹ *Discriminative AI* umumnya digunakan dalam tugas-tugas seperti pendeteksian objek dalam citra digital atau analisis teks untuk mengidentifikasi sentimen. Lahirnya GenAI merupakan fase baru dari kemajuan teknologi kecerdasan buatan, karena GenAI merupakan sistem AI yang mampu menghasilkan data, teks, gambar, video, dan suara baru melalui proses pembelajaran dari data-data yang telah ada.¹⁰

Secara teknis, mekanisme kerja GenAI mengandalkan pemanfaatan *dataset* berskala masif yang terdiri atas jutaan hingga miliaran contoh data digital. Data ini dapat berupa teks, gambar, rekaman suara, hingga metadata perilaku pengguna di internet. Proses pengembangan GenAI dilakukan melalui metode *machine learning* atau lebih khususnya *deep learning*, di mana algoritma model dilatih (*training*

⁸ Bonda Sisephaputra et al., *Generative Artificial Intelligence (GenAI)* (Yogyakarta: PT. Green Pustaka Indonesia, 2024).

⁹ *Ibid.*, hlm. 1.

¹⁰ Anushka Jain, "What is Generative AI," geeksforgeeks, 2025, <https://www.geeksforgeeks.org/artificial-intelligence/what-is-generative-ai/>, diakses pada: 28 Oktober 2025 pukul: 18:05 WIB.

process) untuk mengenali pola, struktur, dan relasi antar data dalam *dataset*.¹¹ Dengan demikian, *dataset* mengambil peran penting dalam proses pengembangan GenAI. Pada pengembangan kecerdasan buatan, *dataset* memegang peran yang sangat fundamental karena menjadi dasar pembelajaran bagi sistem untuk mengenali pola, memahami konteks, dan menghasilkan keputusan atau konten baru.¹²

Dataset merupakan kumpulan data dalam jumlah besar yang dapat berupa teks, gambar, suara, maupun video, yang dikumpulkan dari berbagai sumber baik publik maupun privat. Kualitas dan keragaman *dataset* menentukan tingkat kecerdasan serta keakuratan model AI yang dihasilkan, karena model akan belajar berdasarkan hubungan dan pola yang terkandung di dalam data tersebut. *Dataset* dalam pengembangan kecerdasan buatan umumnya merupakan hasil pengolahan data mentah yang telah melalui proses penyaringan, pembersihan, dan penataan secara sistematis agar dapat digunakan secara efektif dalam melatih model AI.¹³

Pada konteks pengumpulan data dalam *dataset*, terdapat dua metode utama untuk memperoleh data mentah dari internet, yakni *web crawling* dan *web scraping*. *Web crawling* merupakan proses otomatis mengunjungi dan menelusuri situs *web* secara sistematis untuk mengunduh halaman *web* publik dalam skala besar. Proses penelusuran berbagai halaman situs pada *web crawling* dilakukan oleh sebuah program yang bernama *crawler*. Menurut Najork dalam penelitian yang dilakukan

¹¹ Chris Kretz, Christine Fena, dan Ahmad Pratama, "Guide To Generative AI," Stony Brook University, 2024, <https://guides.library.stonybrook.edu/genai/> diakses pada: 28 Oktober 2025, pukul: 18:25 WIB.

¹² *Ibid.*

¹³ *Ibid.*

oleh Farhan, pada proses ini, *crawler* akan melihat suatu halaman *web* serta menelusuri semua *link* yang didapatkan dari halaman tersebut.¹⁴ Salah satu media *raw web data* (sumber utama data mentah) berskala besar yang dimanfaatkan untuk membangun *dataset* adalah Common Crawl.¹⁵

Common Crawl merupakan sumber data mentah terbesar dan paling berpengaruh dalam pengembangan *dataset* untuk GenAI. Sejak tahun 2008, organisasi nirlaba Common Crawl Foundation secara rutin menjalankan proses *web crawling* otomatis setiap bulan dengan mengunjungi miliaran halaman web publik untuk menyalin konten berupa kode HTML, teks, dan metadata yang kemudian disimpan dalam format arsip seperti WARC dan WET agar dapat diakses melalui *Amazon Web Services* (AWS).¹⁶ Peran utama Common Crawl dalam pengembangan *dataset* adalah sebagai infrastruktur data terbuka yang menyediakan data mentah bagi pengembang *Large Language Model* (LLM).¹⁷

Large Language Model adalah model GenAI yang dilatih menggunakan sejumlah besar data teks untuk memahami dan menghasilkan bahasa manusia, contoh GenAI yang merupakan LLM ialah *Generative Pre-Trained Transformer* (GPT), Gemini, Stable Difussion dan lain sebagainya. Dalam pengumpulan data mentah, Common Crawl menggunakan teknik *web crawling*. Salah satu *dataset* yang menggunakan Common Crawl sebagai sumber data mentahnya ialah *dataset*

¹⁴ Farhan Ridho, "Rancang Bangun Aplikasi Web Crawling Untuk Mencari Harga Barang Termurah Dari Berbagai E-Marketplace (Studi Kasus: Tokopedia, Bukalapak, Shopee)" (Universitas Islam Negeri Syarif Hidayatullah, 2020).

¹⁵ Stefan Baack, "A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl," in *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA: ACM, 2024), 2199–2208, <https://doi.org/10.1145/3630106.3659033>.

¹⁶ *Ibid.*, hlm. 2

¹⁷ *Ibid.*, hlm. 2.

LAION-5B. Selain Common Crawl, pengembangan model kecerdasan buatan GenAI juga bergantung pada beragam sumber data mentah lain yang menyediakan materi dasar untuk membangun *dataset*. Sumber-sumber ini mencakup teks ensiklopedik, literatur, media sosial, publikasi ilmiah, hingga repositori kode terbuka.¹⁸

Tahap lanjutan dari proses *web crawling* adalah *web scraping*, yaitu proses mengekstraksi data tertentu dari elemen halaman *web*, seperti *teks*, gambar, atau metadata, dengan menggunakan program otomatis (*scraper*).¹⁹ Menurut Mitchel:

*“In theory, web scraping is the practice of gathering data through any means other than a program interacting with an API (or, obviously, through a human using a web browser). This is most commonly accomplished by writing an automated program that queries a web server, requests data (usually in the form of HTML and other files that compose web pages), and then parses that data to extract needed information.”*²⁰

“Secara teori, web scraping adalah praktik pengumpulan data melalui cara apa pun selain program yang berinteraksi dengan API (atau, tentu saja, melalui manusia yang menggunakan peramban web). Hal ini umumnya dilakukan dengan membuat program otomatis yang mengirimkan permintaan ke server web, meminta data (biasanya dalam bentuk HTML dan berkas lain yang menyusun halaman web), lalu mengurai data tersebut untuk mengekstrak informasi yang dibutuhkan.”

Dengan demikian, *web scraping* merupakan teknik otomatisasi yang memungkinkan pengumpulan informasi dari berbagai situs *web* secara masif dan efisien.²¹

¹⁸ *Ibid.*, hlm. 2.

¹⁹ Yoni Apriyanto, “Apakah Data Scraping Mengancam Bisnis Anda? Analisis Kasus Hukum LinkedIn dan Implikasinya di Indonesia,” Yaplegal.id, 2025, <https://yaplegal.id/blog/apakah-data-scraping-mengancam-bisnis-anda-analisis-kasus-hukum-linkedin-dan-implikasinya-di-indonesia?>, diakses pada: 4 November 2025, pukul: 19:00 WIB.

²⁰ Mitchell Ryan, *Web Scraping with Python*, 2 ed. (O’Reilly Media, 2018).

²¹ *Ibid.*, hlm. 10.

Meskipun data yang diambil sering kali bersumber dari situs publik, aktivitas *web scraping* dapat menimbulkan persoalan hukum ketika informasi yang dikumpulkan mengandung data pribadi yang dapat mengidentifikasi individu. Sebagaimana yang dikemukakan oleh Tzermias, Spiliotopoulos, dan Ioannidis:

*“Even though the data are publicly available, combining and analyzing them at scale can reveal sensitive information about individuals, posing significant privacy risks.”*²²

“Meskipun data tersebut tersedia untuk umum, penggabungan dan analisis data tersebut dalam skala besar dapat mengungkap informasi sensitif mengenai individu, sehingga menimbulkan risiko privasi yang signifikan.”

Pernyataan tersebut menunjukkan bahwa keterbukaan data tidak menghapus sifat pribadi dari suatu informasi, karena ketika dikombinasikan dan dianalisis secara besar-besaran, data yang tampak publik dapat berubah menjadi data sensitif.²³

Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi (selanjutnya disebut sebagai UU PDP) pada Pasal 1 angka 1 menegaskan, bahwa:

“Data Pribadi adalah data tentang orang perseorangan yang teridentifikasi atau dapat diidentifikasi secara tersendiri atau dikombinasi dengan informasi lainnya baik secara langsung maupun tidak langsung melalui sistem elektronik atau non elektronik.”

Merujuk pada Pasal tersebut, esensi dari data pribadi terletak pada informasi yang memiliki kemampuan untuk mengidentifikasi identitas individu atau subjek data tertentu, sekaligus membedakannya dari subjek data lainnya.²⁴

²² Zacharias Tzermias et al., “Privacy Risks From Public Data Sources,” *arXiv:1711.09260*, no. 2 (2017): 1–14, <https://doi.org/https://doi.org/10.48550/arXiv.1711.09260>.

²³ *Ibid.*, hlm. 1.

²⁴ Pasal 1 Angka 1 Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi.

Data pribadi merupakan informasi, baik tunggal maupun gabungan, yang berkaitan dengan seseorang yang telah teridentifikasi atau dapat diidentifikasi keberadaannya.²⁵ Umumnya data yang teridentifikasi dengan individu atau subjek data secara langsung diwujudkan dalam bentuk kode unik yang terdapat pada kartu identitas seperti Kartu Tanda Penduduk (KTP), paspor, Surat Izin Mengemudi (SIM), serta informasi mengenai kondisi fisik, mental, ekonomi, kesehatan, dan data biometrik seperti sidik jari, retina mata, wajah maupun DNA seseorang.²⁶ Sedangkan data yang dapat mengidentifikasi subjek data secara tidak langsung dapat berupa gambar atau potret diri seseorang, *IP address*, data lokasi seseorang, alamat surel, foto profil media sosial, serta data pribadi yang tersedia pada situs publik.²⁷★

Sesuai dengan penggalan Pasal 1 Angka 1 UU PDP:

“...dikombinasi dengan informasi lainnya baik secara langsung maupun tidak langsung melalui sistem elektronik atau non elektronik.”²⁸

Pelindungan data pribadi tidak hanya melindungi data yang diproses melalui sistem non elektronik, melainkan juga melindungi data pribadi yang diproses dengan menggunakan sistem elektronik. Seiring dengan pesatnya kemajuan teknologi digital, data pribadi setiap subjek hukum semakin banyak terekam sistem elektronik, bahkan berpotensi tersebar di berbagai *website* atau situs dan jaringan internet. Oleh karena itu, informasi seperti nama, foto profil, alamat email, lokasi,

²⁵ Sinta Dewi Rosadi, *Pembahasan UU Pelindungan Data Pribadi (UU RI No. 27 Tahun 2022)*, 2 ed. (Jakarta: Sinar Grafika, 2024).

²⁶ *Ibid.*, hlm. 27.

²⁷ *Ibid.*, hlm. 27.

²⁸ Pasal 1 Angka 1 Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi.

hingga metadata digital seperti *IP address* atau *GPS coordinates* yang dikumpulkan melalui *web scraping* dapat dikategorikan sebagai data pribadi.²⁹

Pelindungan terhadap data pribadi mencakup seluruh proses pengumpulan, penyimpanan, dan pemrosesan data, baik melalui sistem elektronik maupun non elektronik. Dalam konteks pengembangan GenAI, proses transformasi data mentah menjadi *dataset* pelatihan yang terstruktur dan berkualitas tinggi, umumnya diawali dengan praktik pengumpulan data menggunakan teknik *web scraping*. Teknik ini memungkinkan sistem untuk secara otomatis mengakses, menyalin, dan mengumpulkan informasi dari berbagai sumber daring guna memperoleh data dalam jumlah besar. Berdasarkan Pasal 16 ayat (1) huruf a UU PDP proses pengumpulan data merupakan bagian dari pemrosesan data pribadi.³⁰ UU PDP menekankan bahwa setiap praktik pemrosesan data pribadi harus berlandaskan prinsip pelindungan data pribadi, di mana pengumpulan data dilakukan secara spesifik, terbatas, serta transparan.³¹

Namun, praktik *web scraping* sering kali dilakukan tanpa mekanisme transparansi dan pemberitahuan kepada pemilik data. Hal ini berpotensi menimbulkan pelanggaran terhadap hak subjek data serta melanggar prinsip transparansi dalam pelindungan data pribadi. Sejalan dengan prinsip dasar UU PDP yang menempatkan subjek data sebagai pihak yang memiliki kendali atas pemrosesan data pribadinya, Rosadi menegaskan bahwa salah satu pilar utama

²⁹ Rosadi, *Pembahasan UU Pelindungan Data Pribadi (UU RI No. 27 Tahun 2022)*, 2 ed.. *Op. cit.*, hlm. 29

³⁰ Pasal 16 ayat (1) huruf a Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi.

³¹ *Ibid.*, Pasal 16 ayat (2) huruf a.

pelindungan data pribadi adalah menjamin kontrol subjek data terhadap proses pemrosesan informasi pribadinya.³²

Sebagai implementasi dari penegasan bahwa subjek data merupakan pihak yang memiliki kendali atas data pribadinya, UU PDP menegaskan hak-hak subjek data melalui Pasal 16 Ayat (2) dengan bunyi:

“Pemrosesan Data Pribadi sebagaimana dimaksud pada ayat (1) dilakukan sesuai dengan prinsip pelindungan data pribadi meliputi:

- a. Pengumpulan Data Pribadi dilakukan secara terbatas dan spesifik, sah secara hukum, dan transparan.”³³

Ketentuan ini menegaskan bahwa subjek data pribadi berhak mendapatkan informasi tentang kejelasan penggunaan data pribadinya, dengan demikian pemrosesan data harus mengedepankan prinsip transparansi. Prinsip transparansi sendiri sudah diterapkan terlebih dahulu dalam *General Data Protection Regulation Uni Eropa* yang selanjutnya disingkat menjadi GDPR pada pasal 5 ayat 1 huruf a yang menegaskan bahwa:

“*Personal data shall be: processed lawfully, fairly, and in a transparent manner in relation to the data subject.*”³⁴

“Data pribadi harus diproses secara sah, adil, dan transparan terhadap subjek data.”

GDPR sudah secara jelas menegaskan bahwa dalam proses pengolahan data pribadi harus memenuhi prinsip transparansi bagi subjek data.

³² Rosadi., Rosadi, *Pembahasan UU Pelindungan Data Pribadi (UU RI No. 27 Tahun 2022)*, *Op. cit.*, hlm. 51.

³³ Pasal 5 Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi.

³⁴ European Union, “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation),” *Official Journal of the European Union*, vol. L119 (Brussels: European Union, 2016), <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

Namun, efektivitas prinsip transparansi masih menghadapi tantangan dalam praktiknya. Penelitian, Wulf dan Seizov mengemukakan bahwa:

*“Current GDPR-mandated disclosures do not meet the expectations and needs of data subjects. Explanations drawn up following the guidelines of the generic formulations of the GDPR differ widely and are often vague, incomplete and lack transparency.”*³⁵

“Pengungkapan informasi yang saat ini diwajibkan oleh GDPR belum memenuhi harapan dan kebutuhan subjek data. Penjelasan yang disusun berdasarkan pedoman rumusan umum dalam GDPR sangat bervariasi dan sering kali tidak jelas, tidak lengkap, serta kurang transparan.”

Penegasan prinsip transparansi terhadap pengelolaan data pribadi dalam GDPR nyatanya tidak memenuhi kebutuhan dan ekspektasi subjek data karena samar, tidak lengkap, dan tidak transparan dalam pengembangan *Artificial Intelligence*.³⁶ Pada hasil penelitiannya, Wulf dan Seizov menegaskan bahwa:

*“Due to massive increases in computing power and big data processing, modern AI algorithms are too complex and opaque to be understood by most data subjects.”*³⁷

“Akibat peningkatan yang sangat pesat dalam daya komputasi dan pemrosesan big data, algoritma AI modern menjadi terlalu rumit dan tidak transparan untuk dipahami oleh sebagian besar subjek data”

Dengan demikian, algoritma AI modern terlalu kompleks dan tidak transparan sehingga sebagian besar subjek data yang datanya digunakan untuk pengembangan AI tidak memahami bagaimana data mereka diproses secara langsung. Dari penelitian yang dilakukan Wulf dan Seizov, disimpulkan bahwa prinsip transparansi

³⁵ Alexander J. Wulf dan Ognyan Seizov, “‘Please understand we cannot provide further information’: evaluating content and transparency of GDPR-mandated AI disclosures,” *AI & Society* 39 (2024): 235–56, <https://link.springer.com/article/10.1007/s00146-022-01424-z>.

³⁶ *Ibid.*, hlm. 235

³⁷ *Ibid.*, hlm. 235

yang diatur oleh GDPR masih belum efektif jika diterapkan pada pengembangan AI.³⁸

Fenomena ini menimbulkan pertanyaan mendasar mengenai bagaimana penerapan transparansi pengumpulan data kepada subjek data pribadi dalam praktik pengumpulan *dataset* melalui teknik pengumpulan data otomatis untuk pengembangan GenAI dapat dijamin berdasarkan Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi. Pertanyaan tersebut semakin relevan ketika dikaitkan dengan akibat hukum yang ditimbulkan apabila subjek data tidak memperoleh transparansi pada pemrosesan data dalam pengembangan GenAI.

Dengan demikian, peneliti bermaksud mendalami persoalan tersebut dengan mengangkat judul, “Transparansi Penggunaan Data Pribadi Dalam Pengumpulan *Dataset* Pada Pengembangan *Generative Artificial Intelligence* Berdasarkan Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 Tentang Pelindungan Data Pribadi”

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka muncul isu permasalahan yang dirumuskan sehingga dapat menjadi rujukan yang runtut dan sistematis dalam pembahasan penelitian ini. Adapun pokok masalah yang diformulasikan, yakni:

³⁸ *Ibid.*, hlm. 236

1. Bagaimana penerapan transparansi data pribadi dalam praktik pengumpulan data untuk pembentukan *dataset* pada pengembangan GenAI berdasarkan UU PDP?
2. Apa akibat hukum pelanggaran prinsip transparansi dalam pengumpulan data pribadi untuk pengembangan GenAI berdasarkan UU PDP?

1.3. Tujuan Penelitian

Adapun kesesuaian berdasarkan latar belakang hingga rumusan masalah yang telah dipaparkan maka rumusan mengenai tujuan dari penelitian ini ialah:

1. Untuk menganalisis pengaturan hukum dan penerapan hak transparansi subjek data pribadi terhadap pengumpulan data secara otomatis melalui teknik *web crawling* dan *web scraping* dalam konteks pembentukan *dataset* pada pengembangan GenAI berdasarkan ketentuan Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi.
2. Untuk mengkaji akibat hukum dari pelanggaran prinsip transparansi dalam pengumpulan data pribadi dalam dataset pada pengembangan GenAI berdasarkan UU PDP.

1.4. Manfaat Penelitian

Adapun manfaat penelitian ini akan memberikan manfaat, baik secara teoritis maupun manfaat secara praktis, yakni:

1.4.1. Manfaat Teoretis

- a. Penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan ilmu hukum, khususnya dalam bidang hukum

perlindungan data pribadi dan hukum teknologi informasi melalui penguatan landasan teoritis mengenai penerapan prinsip transparansi dalam pengelolaan data pribadi yang digunakan untuk pengembangan *Generative Artificial Intelligence* (GenAI).

- b. Pemberian pemahaman konseptual yang lebih mendalam tentang hubungan antara hak subjek data dan kewajiban pengendali data dalam konteks UU Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi (UU PDP).
- c. Kontribusi terhadap pengembangan teori hukum perlindungan data pribadi dan etika kecerdasan buatan, dengan menyoroti pertanggungjawaban perdata akibat pelanggaran transparansi dalam pengolahan data untuk pelatihan algoritma GenAI.

1.4.2. Manfaat Praktis

- a. Bagi Pemerintah

Penelitian ini dapat menjadi pertimbangan dalam penyusunan atau penyempurnaan peraturan pelaksana UU PDP, khususnya yang berkaitan dengan transparansi dan pertanggungjawaban dalam pemrosesan data untuk pengembangan teknologi kecerdasan buatan.

- b. Bagi Pengendali Data dan Pelaku Industri Teknologi

Penelitian ini memberikan pedoman normatif dan etis dalam penerapan prinsip transparansi dan akuntabilitas dalam proses pengumpulan serta pelatihan *dataset* GenAI, guna mencegah pelanggaran hak subjek data.

c. Bagi Masyarakat

Penelitian ini diharapkan dapat meningkatkan kesadaran hukum mengenai hak-hak masyarakat sebagai subjek data dalam pengendalian dan perlindungan data pribadi di era digital.

d. Bagi Akademisi

Penelitian ini dapat menjadi referensi dalam menafsirkan, menerapkan, dan mengembangkan kerangka hukum mengenai tanggung jawab perdata atas pelanggaran perlindungan data pribadi dalam konteks perkembangan teknologi GenAI.

