Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 11, November 2025, Hal. 3425-3436
DOI: https://doi.org/10.52436/1.jpti.1105
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Implementasi Large Language Model dalam Multi-Domain Psikologi: Tinjauan Literatur Sistematis

Mohammad Zakaria Ansor[*1], Dyah Topan Ari Kusuma[2], Jan Everhard Riwurohi[3]

[1,2,3]Magister Ilmu Komputer, Universitas Budi Luhur, Indonesia
Email: [1]2211601519@student.budiluhur.ac.id, [2]2211601600@student.budiluhur.ac.id,
[3]yan.everhard@budiluhur.ac.id

Abstrak

Implementasi *large language models* (LLM) dalam bidang psikologi menyajikan peluang signifikan untuk meningkatkan diagnosis, pengambilan keputusan klinis, dan penelitian medis. Studi ini melakukan tinjauan literatur sistematis untuk mengeksplorasi penelitian-penelitian terkini mengenai aplikasi LLM dalam bidang psikologi. Dengan mengikuti panduan PRISMA, pencarian literatur dilakukan pada database ScienceDirect. Kriteria inklusi dan eksklusi diterapkan untuk mengidentifikasi studi-studi yang relevan. Data yang diekstraksi mencakup tujuan penelitian, metodologi, bidang aplikasi, jenis data yang digunakan, *key findings*, dan hasil. Sebanyak 20 studi dimasukkan setelah proses seleksi. Review ini memberikan gambaran komprehensif mengenai aplikasi LLM dalam bidang psikologi, mengidentifikasi peluang, tantangan, dan arah penelitian masa depan yang bermanfaat bagi peneliti, praktisi, dan pembuat kebijakan. Temuan ini menunjukkan bahwa integrasi LLMs dalam praktik psikologi memiliki potensi transformatif untuk meningkatkan kualitas dan aksesibilitas layanan kesehatan mental, namun memerlukan pengembangan framework etis dan regulasi yang komprehensif untuk memastikan implementasi yang aman dan efektif.

Kata kunci: *Large Language Models, Literature Review, Psikologi*

Implementation of Large Language Models in Multi-Domain Pshycology : A Systematic Literature Review

Abstract

The implementation of large language models (LLMs) in psychology presents significant opportunities for enhancing diagnostic procedures, clinical decision-making, and medical research. This study conducted a systematic literature review to explore contemporary research regarding LLM applications in the field of psychology. Following PRISMA guidelines, a literature search was performed using the ScienceDirect database. Inclusion and exclusion criteria were applied to identify relevant studies. Extracted data encompassed research objectives, methodologies, application domains, data types utilized, key findings, and outcomes. A total of 20 studies were included following the selection process. This review provides a comprehensive overview of LLM applications in psychology, identifying opportunities, challenges, and future research directions that will benefit researchers, practitioners, and policymakers. The findings indicate that the integration of LLMs into psychological practice demonstrates transformative potential for improving the quality and accessibility of mental health services; however, it necessitates the development of comprehensive ethical frameworks and regulatory standards to ensure safe and effective implementation.

Keywords: *Large Language Models, Literature Review, Psychology*

1. PENDAHULUAN

Perkembangan *Large Language Models* (LLM) telah mengalami evolusi yang signifikan dan semakin terintegrasi dalam kehidupan sehari-hari manusia. Model-model ini dapat diakses melalui *interface* yang mudah digunakan seperti ChatGPT untuk berbagai keperluan, mulai dari pencarian informasi, bantuan akademik, hingga layanan pelanggan. Kemampuan LLM dalam memproses dan menghasilkan bahasa natural yang mudah dipahami manusia semakin meningkat seiring dengan bertambahnya jumlah parameter dan data training yang digunakan. Bahkan terdapat bukti empiris yang menunjukkan bahwa teks yang dihasilkan oleh LLM sering kali dipersepsikan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

sebagai tulisan manusia asli dan sering kali model LLM juga mampu mendemonstrasikan kemampuan *reasoning* yang canggih dan strategi negosiasi tingkat tinggi[1].

Kemajuan pesat dalam bidang *generative artificial intelligence* telah menciptakan antusiasme sekaligus kekhawatiran di kalangan komunitas ilmiah. Sejak peluncuran ChatGPT pada November 2022, volume penelitian yang membahas potensi manfaat dan risiko penggunaan AI dalam riset mengalami pertumbuhan yang eksponensial. ChatGPT merupakan *web interface* yang memungkinkan pengguna berkomunikasi dengan jenis LLM yang disebut *'generative pre-trained transformer*' atau 'GPT', di mana istilah *generative* merujuk pada kemampuan model dalam menghasilkan teks, *pre-trained* menandakan bahwa model telah mengalami *supervised* dan *unsupervised training* menggunakan sumber data online yang masif, sedangkan transformer mengacu pada kemampuan model dalam menganalisis detail spesifik dalam struktur dan konten teks untuk menghasilkan output[2]. Aksesibilitas AI kini telah meluas ke siapa saja yang memiliki koneksi internet, tidak terbatas pada individu yang memiliki pengetahuan programing, sehingga membuka peluang besar bagi penerapan LLM dalam berbagai disiplin ilmu termasuk psikologi.

Evolusi berkelanjutan dari kapabilitas LLM memunculkan kebutuhan untuk memahami model-model tersebut secara lebih mendalam, dan sejumlah penelitian mulai mengkaji LLM dari perspektif *behavioral*. Hasil penelitian menunjukkan bahwa bias sosial yang terdapat dalam data training dapat terinternalisasi dalam *word embeddings* sehingga LLM terkadang menunjukkan ketidakselarasan dengan nilai-nilai manusia, yang merupakan tantangan yang belum terselesaikan dan memiliki implikasi terhadap keamanan atas AI[3]. Kompleksitas ini menjadi semakin relevan dalam konteks psikologi, di mana pemahaman terhadap perilaku dan proses mental manusia menjadi kunci dalam pengembangan aplikasi LLM yang efektif. Upaya untuk memahami asal-usul dan solusi terhadap perilaku tersebut menghadapi tantangan signifikan yaitu miliaran parameter yang terkandung dalam model-model ini secara substansial memperumit *assessment* analitik terhadap mekanisme internal model, sehingga diperlukan pendekatan alternatif untuk mempelajari implementasi LLM dalam domain psikologis.

Dalam konteks implementasi LLM di bidang psikologi, perdebatan filosofis fundamental tentang sifat pemahaman bahasa dan kognisi menjadi semakin krusial. Kritik yang disampaikan oleh tokoh-tokoh seperti Noam Chomsky terhadap LLM modern menyoroti keterbatasan sistem yang bergantung pada pendekatan *data-driven* dibandingkan dengan sistem berbasis aturan yang mencerminkan kemampuan kognitif manusia yang sesungguhnya[4]. Perdebatan ini memiliki implikasi langsung terhadap aplikasi LLM dalam psikologi, khususnya dalam memahami apakah model-model ini dapat secara otentik mereplikasi atau memfasilitasi proses psikologis manusia. Oleh karena itu, literature review mengenai implementasi LLM di bidang psikologi menjadi sangat penting untuk mengidentifikasi sejauh mana teknologi ini dapat berkontribusi pada penelitian psikologis, terapi, assessment, dan intervensi, serta untuk mengeksplorasi tantangan etis dan metodologis yang muncul dalam penerapannya.

Kemampuan dari LLM dalam menjawab pertanyaan sangat ditentukan oleh *prompt* yang dimasukkan oleh pengguna dan respon yang diberikan masih belum bisa melihat konteks dari pertanyaan psikologi yang diajukan sehingga sering kali respon yang diberikan hanyalah respon yang bersifat umum[5]. Perkembangan Large Language Model yang sangat cepat masih membuka peluang untk dilakukan penelitian terhadap batasan tanggung jawab dalam skenario penggunaannya dalam bidang psikologi[6]. Selain itu perkembangan LLM yang begitu pesat membuka peluang yang sangat besar untuk merubah ilmu psikologi dengan meningkatkan akurasi hasil assessment dengan mengakomodasi setiap kata dan ucapan dari penggunanya[7].

Meskipun ketertarikan terhadap implementasi LLM dalam psikologi terus meningkat, tinjauan literatur sistematis (SLR) yang komprehensif pada area ini masih relatif terbatas. Hal ini disebabkan oleh beberapa faktor yaitu perkembangan LLM yang sangat cepat sehingga banyak literatur yang tersebar dan belum terintegrasi secara sistematis dan kompleksitas interdisipliner yang melibatkan ilmu komputer dan ilmu psikologi.

2. METODE PENELITIAN

Studi ini menggunakan metodologi *systematic literature review*, mengikuti panduan *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA)[4].

Pencarian literatur dibatasi hanya dilakukan pada database ScienceDirect menggunakan kata kunci "*large language model*", kemudian pemilihan jurnal yang bersifat *open access*, kemudian dipilih lagi khusus *research articles*, dan berikutnya dipilih lagi yang berkaitan dengan bidang psikologi. Pencarian dilakukan pada judul artikel, abstrak, dan kata kunci untuk memastikan cakupan yang komprehensif. Rentang waktu untuk seleksi artikel adalah antara tahun 2023 sampai dengan 2025.

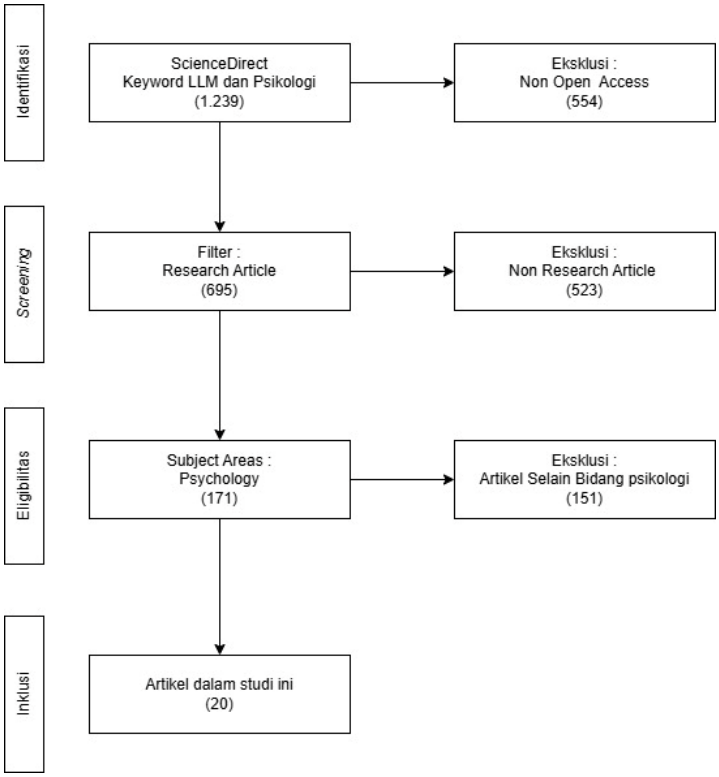
Proses seleksi studi dilakukan pada tanggal 26 Mei 2025 dengan melibatkan beberapa tahapan, seperti yang diilustrasikan dalam Gambar 1. Pada awalnya, artikel disaring berdasarkan judul dan abstrak untuk mengecualikan studi yang tidak relevan. Artikel yang tersisa kemudian dinilai lebih lanjut berdasarkan kriteria inklusi dan eksklusi yang telah ditetapkan:

p-ISSN: 2775-4227
e-ISSN: 2775-4219

1. Studi harus membahas aplikasi LLM dalam bidang psikologi.
2. Artikel harus bersifat *open-access*.
3. Harus merupakan *research paper* dengan metodologi yang jelas dan hasil yang dapat diukur.

Setelah artikel berhasil dikumpulkan, ekstraksi data dilakukan untuk menjawab pertanyaan penelitian berikut:

1. Apa tujuan penerapan LLM dalam bidang psikologi? Pertanyaan penelitian ini berisi luaran penelitian yang diharapkan dari para peneliti.
2. Apa metodologi penelitiannya? Pertanyaan penelitian ini berisi metode penelitian yang dilakukan selama penelitian.
3. Di bidang psikologi mana penelitian dilakukan? Pertanyaan penelitian ini berisi area atau bidang kesehatan yang dipelajari.
4. Data apa yang digunakan? Pertanyaan penelitian ini berisi data yang digunakan dalam penelitian.
5. Apa hasil penelitiannya? Pertanyaan penelitian ini berisi temuan penelitian dan kesimpulan.
6. Data yang diekstraksi diorganisir menggunakan *spreadsheet* dan disintesis dalam *research paper* akhir.



Gambar 1. Metode Pengumpulan Artikel

3. HASIL DAN PEMBAHASAN

Setelah melakukan proses seleksi studi sesuai dengan kriteria inklusi dan eksklusi, total 20 studi yang relevan dengan topik aplikasi LLM dalam bidang psikologi berhasil diidentifikasi. Hasil dan pembahasan dari studi-studi tersebut akan disajikan berdasarkan pertanyaan penelitian yang telah ditentukan sebelumnya.

3.1. Tujuan Pemanfaatan LLM di Bidang Psikologi

Berdasarkan analisis terhadap berbagai tujuan penelitian implementasi *Large Language Models* (LLM) dalam bidang psikologi, dapat diidentifikasi beberapa kluster tematik yang menunjukkan arah pengembangan teknologi ini dalam konteks psikologi klinis dan teoritis.

3.1.1. Kluster Analisis Psikologi Klinis dan Diagnostik

Sejumlah penelitian berfokus pada pemanfaatan LLM untuk mendukung proses diagnostik dan analisis kondisi psikologis pasien. Penelitian ini mengeksplorasi kemampuan LLM dalam menganalisis teks naratif tidak

p-ISSN: 2775-4227
e-ISSN: 2775-4219

terstruktur dari *electronic health records* (EHR) untuk mengekstrak wawasan mendalam dari catatan personal pasien yang sebelumnya sulit dianalisis menggunakan metode konvensional[2]. Dalam konteks yang serupa, LLM juga dimanfaatkan untuk menganalisis gejala yang relevan dengan *Research Domain Criteria* (RDoC) pada pasien dengan gangguan psikiatri, dengan fokus evaluasi kemampuan model dalam mengidentifikasi pola gejala lintas domain kognitif, emosional, dan perilaku guna mendukung diagnosis dan pengembangan intervensi klinis[8]. Selain itu, terdapat upaya pengembangan sistem deteksi risiko bunuh diri melalui analisis narasi individu menggunakan LLM untuk membantu profesional kesehatan mental dalam mendeteksi risiko secara lebih akurat dan efisien[9]. Penelitian lainnya menyelidiki kemampuan LLM dalam mengidentifikasi dan mengukur dimensi gangguan pemikiran pada pasien dengan kondisi psikiatri, khususnya dalam mengenali pola abnormal dalam *discourse* bahasa yang merupakan karakteristik gangguan pemikiran formal[10].

3.1.2. Kluster Aplikasi Terapi dan Intervensi Psikologis

Beberapa penelitian mengeksplorasi implementasi LLM dalam konteks terapi dan intervensi psikologis langsung. Salah satu fokus utama adalah evaluasi efektivitas Amanda, sebuah *voice-based* LLM, dalam memahami, merespons, dan memberikan solusi terhadap berbagai skenario percakapan manusia, dengan penekanan pada kemampuan memberikan respons yang relevan, akurat, dan kontekstual di berbagai bidang aplikasi termasuk layanan kesehatan[11]. Penelitian lainnya menguji *feasibility* penggunaan LLM, khususnya ChatGPT-4, untuk menghasilkan *exposure hierarchies* yang tepat dalam pengobatan *Obsessive-Compulsive Disorder* (OCD), dengan evaluasi dampak berbagai fitur klinis dan demografis terhadap output yang dihasilkan serta perbandingan performa dengan terapis ahli tingkat doctoral[12].

3.1.3. Kluster Analisis Kognitif dan Kemampuan Pemrosesan

Penelitian dalam kluster ini mengeksplorasi kemampuan kognitif inherent dalam LLM dan kaitannya dengan pemrosesan kognitif manusia. Terdapat upaya identifikasi kemampuan kognitif yang saling terkait dalam LLM dengan fokus pada perilaku yang menyerupai pemrosesan kognitif manusia, serta pemahaman sejauh mana kemampuan tersebut muncul secara spontan dari arsitektur dan training model[13]. Penelitian komplementer mengeksplorasi kemampuan LLM dalam memahami dan mereplikasi kemampuan penalaran induktif manusia, dengan identifikasi sejauh mana LLM dapat meniru pola penalaran manusia dalam berbagai konteks serta evaluasi potensi dan keterbatasan model dalam tugas yang melibatkan generalisasi dan inferensi[14]. Selain itu, terdapat evaluasi komprehensif mengenai kemampuan LLM dalam memahami, memproses, dan menghasilkan teks secara individual maupun kolektif untuk mengidentifikasi potensi dan keterbatasan model dalam berbagai konteks aplikasi[15].

3.1.4. Kluster Bias, Etika, dan Moral dalam LLM

Aspek etis dan moral dalam implementasi LLM menjadi fokus penting dengan beberapa penelitian yang menganalisis interaksi LLM dengan nilai-nilai moral dalam berbagai konteks, serta identifikasi sejauh mana model mampu memahami, merepresentasikan, dan mematuhi prinsip moral yang kompleks, termasuk pengukuran potensi risiko dan manfaat penerapan LLM dalam pengambilan keputusan berbasis moral[16]. Penelitian terkait mengeksplorasi bagaimana interaksi dengan LLM dapat dipengaruhi oleh bias atau prasangka, baik yang berasal dari model maupun pengguna, dengan fokus pemahaman sejauh mana bias muncul dan implikasinya terhadap penggunaan LLM di berbagai konteks[17].

3.1.5. Kluster Analisis Linguistik dan Kepribadian

Beberapa penelitian berfokus pada kemampuan LLM dalam analisis linguistik dan *assessment* kepribadian. Penelitian ini meliputi pengembangan metode identifikasi dan pemodelan perbedaan bahasa terkait kepribadian dalam teks menggunakan *small semantic vector subspaces* dari LLM untuk mengekstrak perbedaan linguistik antara individu dengan skor tinggi dan rendah pada dimensi kepribadian Big Five[18]. Penelitian komplementer mengeksplorasi kemampuan ChatGPT dalam mengenali dan memahami impoliteness dalam teks, serta evaluasi akurasi dan batasan model dalam memberikan respons yang sesuai terhadap teks yang mengandung ketidaksopanan[19].

3.1.6. Kluster Studi Fenomenologis dan Filosofis

Terdapat pendekatan filosofis dan fenomenologis dalam memahami LLM, termasuk eksplorasi fenomena halusinasi dalam LLM melalui lensa teori "*cognitive phantom*" dengan fokus pemahaman mekanisme produksi konten yang salah namun tampak meyakinkan, serta identifikasi mekanisme di balik halusinasi dan kategorisasi

p-ISSN: 2775-4227
e-ISSN: 2775-4219

jenisnya[1]. Penelitian filosofis lainnya mengeksplorasi LLM melalui kerangka teoretis filosofi Martin Heidegger dan psikoanalisis Jacques Lacan untuk menguji sifat ontologis LLM serta implikasi kehadiran teknologi AI generatif dalam konteks eksistensial dan subjektif manusia[3]. Penelitian dalam kluster ini juga mengeksplorasi bagaimana interaksi dengan LLM memengaruhi persepsi diri individu, termasuk pandangan mereka terhadap kemampuan, kepribadian, dan identitas setelah berinteraksi dengan teknologi berbasis LLM[20].

3.1.7. Kluster Pengembangan dan Optimalisasi Teknis

Beberapa penelitian berfokus pada aspek teknis pengembangan LLM, termasuk evaluasi efektivitas LLM dengan berbagai ukuran dalam menyelesaikan tugas spesifik untuk mengidentifikasi apakah peningkatan ukuran model berbanding lurus dengan peningkatan performa, serta pemahaman batasan dan *trade-off* dalam pengembangan model yang lebih besar[21]. Penelitian lainnya mengeksplorasi potensi dan tantangan penggunaan LLM dalam mendukung pengembangan *generative artificial intelligence* di berbagai bidang[22], serta eksplorasi kemampuan LLM dalam menganalisis dan menghasilkan *case reports* dengan fokus peningkatan efisiensi, akurasi, dan kualitas penyusunan laporan berbasis teks[23]. Terdapat pula penelitian yang mengeksplorasi penggunaan LLM untuk menganalisis wawasan dari perspektif pemberi kerja dengan identifikasi pola, tren, dan wawasan strategis terkait tenaga kerja[24].

3.2. Metodologi Penelitian dalam Implementasi LLM

3.2.1. Pendekatan Analisis Teoretis dan Evaluasi Performa Model

Penelitian terkait implementasi *Large Language Models* (LLM) dalam bidang psikologi menggunakan kombinasi analisis teoretis dan pengujian empiris yang komprehensif. Pendekatan ini melibatkan kajian mendalam terhadap arsitektur LLM dan mekanisme kerja model-model seperti GPT-3, GPT-3.5, GPT-4, dan PaLM melalui eksperimen sistematis untuk mengevaluasi berbagai fenomena, termasuk halusinasi dan efektivitas teknik mitigasi seperti *chain-of-thought* (CoT) *reasoning*[1]. Metodologi serupa diterapkan dalam evaluasi performa LLM pada tugas-tugas berbasis bahasa, di mana model diuji melalui simulasi dan eksperimen berbasis data untuk mengukur kemampuan dalam menghasilkan respons yang relevan dan berkualitas tinggi pada pemahaman teks, penyelesaian masalah, dan kolaborasi antar-model[15]. Pendekatan eksperimental juga digunakan untuk mengevaluasi kemampuan LLM dalam memahami, menghasilkan, dan memproses bahasa alami secara kontekstual[22].

3.2.2. Metodologi Analisis dan Klasifikasi Teks Klinis

Beberapa penelitian mengadopsi pendekatan *machine learning* dengan memanfaatkan LLM untuk memproses dan menganalisis data teks dari laporan klinis dan wawancara pasien. Metodologi ini mencakup pelatihan model untuk mengenali dan mengklasifikasikan gejala berdasarkan kerangka kerja *Research Domain Criteria* (RDoC), dengan validasi melalui perbandingan hasil analisis model dengan penilaian manual oleh ahli psikiatri[8]. Implementasi serupa terlihat pada pengujian tiga LLM berbeda (GPT-3.5, GPT-4, dan Claude) untuk mengklasifikasikan 500 narasi klinis pasien ke dalam empat level risiko bunuh diri menggunakan prompt terstruktur, dengan evaluasi kinerja berdasarkan metrik akurasi, *recall, precision, F1-score*, dan korelasi Spearman yang dibandingkan dengan penilaian klinis dan model pembelajaran mesin tradisional seperti BERT, XGBoost, dan Random Forest[9].

3.2.3. Pendekatan Analisis Tematik dan Naratif

Metodologi analisis tematik menggunakan LLM seperti GPT untuk menganalisis teks naratif melalui proses pengkodean otomatis, analisis tematik, dan evaluasi hasil oleh peneliti untuk menilai akurasi dan relevansi interpretasi yang dihasilkan model[2]. Pendekatan ini diperluas dalam konteks pembuatan laporan kasus, di mana LLM diuji untuk menghasilkan laporan berdasarkan data yang diberikan dengan evaluasi kemampuan memahami konteks, menyusun narasi yang relevan, dan memberikan informasi akurat melalui perbandingan dengan laporan yang disusun manusia[23].

3.2.4. Metodologi Evaluasi Interaksi dan Bias

Penelitian interaksi pengguna LLM menggunakan desain eksperimental dengan skenario interaksi yang dirancang khusus untuk memunculkan bias eksplisit maupun implisit, diikuti analisis respons model dan pengamatan pola-pola bias berdasarkan parameter tertentu[17]. Metodologi serupa diterapkan dalam eksperimen berbasis interaksi langsung dengan ChatGPT, di mana peneliti menyusun skenario komunikasi mencakup teks

p-ISSN: 2775-4227
e-ISSN: 2775-4219

sopan, netral, dan tidak sopan, kemudian menganalisis respons model secara kualitatif dan kuantitatif untuk mengevaluasi kemampuan deteksi dan respons terhadap ketidaksopanan[19].

3.2.5. Pendekatan Evaluasi Moral dan Etika

Evaluasi aspek moral LLM dilakukan melalui serangkaian skenario moral yang dirancang untuk mengukur respons terhadap dilema etika, dengan evaluasi menggunakan kerangka kerja analisis kuantitatif dan kualitatif terhadap output model, serta perbandingan performa beberapa LLM untuk mengidentifikasi pola dan perbedaan dalam pengambilan keputusan moral[16].

3.2.6. Metodologi Berbasis Kerangka Filosofis

Pendekatan interdisipliner mengintegrasikan filsafat Heidegger dan psikoanalisis Lacan sebagai kerangka konseptual untuk menganalisis bagaimana LLM seperti GPT memproses bahasa dan menghasilkan teks, dengan mengaitkan proses tersebut pada konsep "kekosongan" dalam teori Heidegger dan Lacan melalui eksperimen kecil untuk menguji respons model terhadap pertanyaan atau perintah berkaitan konsep filosofis[3].

3.2.7. Pendekatan Simulasi dan Interaksi Langsung

Metodologi pengujian langsung diterapkan melalui simulasi percakapan berbasis suara dengan model seperti Amanda, menggunakan skenario percakapan yang telah dirancang dan interaksi langsung dengan pengguna manusia, dengan evaluasi berdasarkan metrik tingkat akurasi respons, relevansi konteks, dan kepuasan pengguna[11]. Pendekatan serupa melibatkan partisipan yang berinteraksi dengan LLM dalam berbagai skenario, diikuti evaluasi persepsi diri melalui survei dan wawancara untuk mengukur perubahan pandangan terhadap diri sendiri[20].

3.2.8. Metodologi Evaluasi Kemampuan Kognitif dan Penalaran

Pengujian kemampuan kognitif LLM dilakukan melalui serangkaian tugas yang dirancang untuk mengevaluasi kemampuan spesifik seperti penalaran, pemecahan masalah, pemahaman konteks, dan generalisasi menggunakan prompt yang dirancang untuk memancing respons yang mencerminkan kemampuan kognitif tersebut[13]. Metodologi eksperimental dengan LLM sebagai subjek utama diterapkan untuk menjalankan tugas penalaran induktif, dengan evaluasi melalui eksperimen berbasis teks yang dibandingkan dengan data perilaku manusia dari studi sebelumnya menggunakan analisis kuantitatif untuk mengukur tingkat akurasi dan kesesuaian respons terhadap pola penalaran manusia[14].

3.2.9. Pendekatan Analisis Komputasional Spesialisasi

Metodologi analisis komputasional menggunakan *Principal Component Analysis* (PCA) pada *embedding* kalimat dari teks individu dengan skor kepribadian berbeda untuk mengidentifikasi *subspace* semantik yang menangkap variasi bahasa terkait kepribadian melalui pendekatan Pe-LLM (Personality-LLM) yang terdiri dari ekstraksi *subspace* semantik relevan dengan dimensi kepribadian dan prediksi skor kepribadian dari teks baru[18]. Pendekatan serupa diterapkan menggunakan GPT-4 dan Claude untuk mengevaluasi transkrip wawancara klinis berdasarkan *Thought Disorder Index* (TDI), dengan pelatihan model menggunakan contoh gangguan pemikiran dan penilaian transkrip berdasarkan 23 kategori gangguan pemikiran dengan empat tingkat keparahan, kemudian membandingkan hasil dengan penilaian psikolog klinis berpengalaman[10].

3.2.10. Metodologi Perbandingan dan Optimisasi Model

Penelitian komparatif dilakukan untuk membandingkan performa beberapa LLM dengan ukuran berbeda pada tugas linguistik dan non-linguistik melalui analisis akurasi, efisiensi, dan kemampuan generalisasi terhadap data yang diberikan, serta mengkaji dampak ukuran model terhadap konsumsi sumber daya komputasi[21]. LLM juga digunakan untuk memproses dan menganalisis dokumen terkait pemberi kerja seperti laporan tahunan, survei tenaga kerja, dan deskripsi pekerjaan dengan evaluasi kemampuan identifikasi pola dan generasi wawasan yang relevan[24].

3.2.11. Metodologi Studi *Feasibility* Klinis

Desain studi *feasibility* dengan simulasi kasus pasien menggunakan ChatGPT (GPT-4) untuk menghasilkan 10-*item exposure hierarchies* dengan variasi *prompt* sistematis berdasarkan dimensi OCD *subtype*, kompleksitas gejala, level detail, usia, dan gender pasien. Evaluasi dilakukan melalui *initial review* oleh staf psikolog untuk

p-ISSN: 2775-4227
e-ISSN: 2775-4219

menguji kelengkapan dan penggabungan informasi yang diinput, diikuti *blinded review* oleh tenaga ahli klinis untuk menguji kelayakan, kekhususan, variabilitas, keamanan/etik, dan kegunaan secara umum, dengan pembanding berupa hierarki yang dibuat tenaga klinis ahli[12].

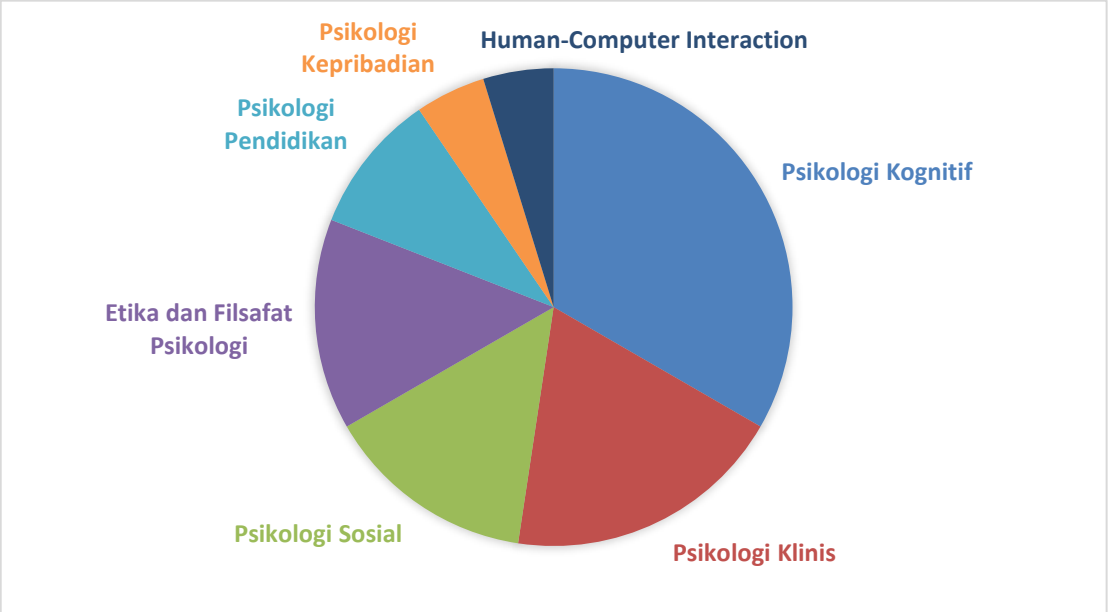
3.3. Bidang Studi Psikologi

Distribusi bidang studi penelitian menunjukkan bahwa implementasi LLM dalam psikologi paling dominan pada bidang kognitif (33.3%) dan klinis (19.0%), yang mencerminkan dua area aplikasi utama: pemahaman proses mental dan aplikasi terapeutik. Bidang etika dan sosial juga mendapat perhatian signifikan (masing-masing 14.3%), menunjukkan kesadaran akan dampak sosial dan moral dari teknologi AI. Sementara itu, aplikasi dalam pendidikan, kepribadian, dan HCI masih relatif terbatas, menunjukkan potensi pengembangan lebih lanjut di masa depan. Adanya penelitian dengan klasifikasi ganda[19] menunjukkan sifat interdisipliner dari penelitian LLM dalam psikologi, di mana satu penelitian dapat mencakup beberapa dimensi psikologis secara bersamaan. Penelitian[15] juga memiliki karakteristik interdisipliner dengan menggabungkan *natural language processing* (aspek kognitif) dan ilmu sosial komputasional (aspek sosial), namun diklasifikasikan primer ke Psikologi Kognitif karena fokus utamanya pada pemrosesan bahasa sebagai fungsi kognitif.

Bidang psikologi kognitif mendominasi penelitian LLM dalam psikologi dengan 7 penelitian. Penelitian-penelitian dalam kategori ini mengeksplorasi bagaimana teknologi AI dapat memahami dan memproses bahasa manusia, yang merupakan fungsi kognitif fundamental. Penelitian[1] membahas kecerdasan buatan dan pemrosesan bahasa alami, penelitian[15] mengintegrasikan *natural language processing* dengan ilmu sosial komputasional, penelitian[11] menggabungkan pemrosesan bahasa alami dengan teknologi pengenalan suara, penelitian[23] fokus pada pemrosesan bahasa alami dan aplikasi kecerdasan buatan, penelitian[13] menghubungkan AI dengan ilmu kognitif dan NLP, penelitian[14] mengkaji ilmu kognitif bersama AI dan linguistik komputasional, dan penelitian[21] fokus pada NLP dan kecerdasan buatan.

Tabel 1. Distribusi Penelitian per Bidang Studi

Bidang Psikologi	Jml Penelitian	Penelitian
Psikologi Kognitif	7	[1], [11], [13], [14], [15], [21], [23]
Psikologi Klinis	4	[2], [9], [10], [12]
Psikologi Sosial	3	[15], [17], [19]
Etika dan Filsafat Psikologi	3	[3], [16], [19]
Psikologi Pendidikan	2	[22], [24]
Psikologi Kepribadian	1	[18]
Human-Computer Interaction	1	[20]



Gambar 2. Diagram Penelitian Bidang Psikologi

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Kelompok kedua terbesar adalah bidang psikologi klinis mencakup aplikasi LLM dalam konteks klinis dan kesehatan mental. Penelitian[2] menganalisis pengalaman pasien dan pemanfaatan teknologi AI dalam pelayanan kesehatan, penelitian[9] menerapkan NLP untuk kesehatan mental dengan pendekatan *data science*, penelitian[18] mengembangkan psikiatri komputasional yang mengintegrasikan psikologi klinis dengan AI, dan penelitian[12] secara spesifik memfokuskan pada psikologi klinis. Bidang ini menunjukkan potensi besar LLM dalam diagnosis, terapi, dan pemahaman gangguan mental.

Dalam bidang psikologi sosial, tiga penelitian mengeksplorasi dimensi sosial dari implementasi LLM. Penelitian[15] menggabungkan *natural language processing* dengan ilmu sosial komputasional, penelitian[19] mengkaji interaksi manusia-mesin dari perspektif linguistik komputasional, dan penelitian[17] membahas studi sosial dalam konteks etika teknologi AI. Pengelompokan ini menunjukkan bagaimana teknologi AI mempengaruhi interaksi sosial dan dinamika kelompok.

Bidang etika dan filsafat psikologi mencerminkan refleksi mendalam tentang implikasi moral dan filosofis dari LLM. Penelitian[3] mengintegrasikan filsafat teknologi, etika AI, dan psikoanalisis, penelitian[16] membahas kecerdasan buatan dari perspektif etika dan filsafat moral, sementara penelitian[19] juga mencakup aspek etika kecerdasan buatan. Pengelompokan ini penting karena menunjukkan kesadaran akan tanggung jawab etis dalam pengembangan teknologi AI.

Dalam bidang psikologi Pendidikan, Dua penelitian mengeksplorasi aplikasi LLM dalam konteks pendidikan dan pengembangan sumber daya manusia. Penelitian[22] membahas implementasi LLM dalam pendidikan, kesehatan, bisnis, dan teknologi kreatif, sedangkan penelitian[24] fokus pada *natural language processing* dalam manajemen SDM. Bidang ini menunjukkan potensi LLM dalam meningkatkan proses pembelajaran dan pengembangan keterampilan.

Bidang psikologi kepribadian, Penelitian[18] secara spesifik mengintegrasikan psikologi kepribadian dengan pemrosesan bahasa alami dan kecerdasan buatan, menunjukkan bagaimana LLM dapat memahami dan menganalisis karakteristik kepribadian individu melalui analisis bahasa.

Terakhir Penelitian[20] dikelompokkan dalam bidang *human-computer interaction* karena menggabungkan ilmu komputer, psikologi, dan studi interaksi manusia dengan komputer, yang merupakan bidang interdisipliner yang mengkaji bagaimana manusia berinteraksi dengan teknologi komputer, termasuk sistem AI dan LLM

3.4. Data Penelitian

Beberapa penelitian telah mengeksplorasi kemampuan kognitif *Large Language Models* (LLM) dengan menggunakan dataset yang bervariasi dalam kompleksitas dan cakupan. Dataset pengetahuan umum dengan format trivia pilihan ganda telah digunakan untuk menguji empat model berbeda (GPT-3.5, GPT-4, Claude, PaLM2) menggunakan 50 pertanyaan yang menghasilkan 25.000 respons model untuk dibandingkan dengan data dari 889 partisipan[15]. Sementara itu, evaluasi yang lebih komprehensif dilakukan menggunakan 250 pertanyaan yang dikategorikan ke dalam lima domain utama yaitu faktual, analitis, kreatif, temporal, dan responsif, dengan tambahan 75 pertanyaan lanjutan untuk menguji kemampuan mempertahankan konteks percakapan[11]. Kemampuan penalaran induktif juga dievaluasi menggunakan tugas-tugas yang didasarkan pada eksperimen psikologi kognitif sebelumnya dengan memanfaatkan korpus pelatihan yang beragam dari berbagai domain literatur dan penelitian ilmiah[14].

Implementasi LLM dalam konteks kesehatan mental telah diteliti melalui analisis naratif pengalaman pasien yang dikumpulkan dari berbagai sumber klinis. Seratus naratif dari pasien rawat inap di rumah sakit komunitas kecil Amerika Serikat dianalisis dengan rata-rata panjang 32,8 kata per naratif untuk memahami pengalaman perawatan[2]. Dalam skala yang lebih besar, 500 narasi klinis diekstrak dari catatan elektronik pasien di Massachusetts General Brigham, mencakup catatan psikiatri rawat jalan, rawat inap, layanan darurat, dan berbagai spesialis, dengan label level risiko bunuh diri (1-4) yang diberikan oleh tiga klinisi berpengalaman sebagai *gold standard*[9]. Data klinis juga diperoleh melalui transkrip wawancara pasien, catatan medis elektronik, dan laporan klinis yang dianonimkan untuk mengekstraksi informasi gejala *Research Domain Criteria* (RDoC) menggunakan LLM[8].

Penelitian khusus dalam bidang psikologi klinis menggunakan 36 transkrip wawancara dari tiga kelompok subjek: pasien skizofrenia (n=12), pasien gangguan bipolar (n=12), dan kontrol sehat (n=12). Transkrip ini berasal dari *Rorschach inkblot test* yang dirancang untuk memunculkan respons pemikiran kompleks dan mendeteksi pola gangguan pemikiran, dengan penilaian manual menggunakan *Thought Disorder Index* (TDI) oleh psikolog klinis[10]. Evaluasi *diagnostic hierarchy* juga dilakukan menggunakan 72 *prompt* yang mewakili kombinasi dimensi yang divariasikan, menghasilkan 55 respons lengkap dan 15 respons parsial dari ChatGPT, yang dibandingkan dengan 18 hierarki yang dibuat oleh tenaga klinis ahli[12].

Penelitian tentang bias sosial dalam LLM menggunakan 2.400 prompt yang dirancang khusus mengandung prasangka terhadap 26 kelompok sosial berbeda, mencakup kategori ras, etnis, agama, gender, orientasi seksual,

p-ISSN: 2775-4227
e-ISSN: 2775-4219

kelas sosial, dan disabilitas. Prompt ini dibuat dalam bentuk prasangka eksplisit dan implisit, diujikan pada lima LLM berbeda, menghasilkan total 12.000 interaksi untuk analisis[17]. Aspek etika moral juga dievaluasi menggunakan skenario moral yang dirancang khusus, termasuk dilema etika klasik seperti *trolley problem* dan situasi moral kontekstual yang lebih kompleks, dengan memanfaatkan data pelatihan LLM dari berbagai sumber seperti buku, penelitian, dan forum online[16].

Studi tentang ketidaksopanan dalam komunikasi menggunakan 33 percakapan pendek berbahasa Inggris yang dirancang untuk menunjukkan enam kategori ketidaksopanan: penghinaan langsung, penghinaan tidak langsung, penolakan bantuan, kritik, perintah, dan sarkasme, yang dibuat berdasarkan teori ketidaksopanan untuk mewakili situasi sehari-hari[19]. Penelitian *cross-cultural* dilakukan dengan melibatkan 643 subjek mahasiswa Spanyol dengan rata-rata usia 19,50 tahun yang merupakan *native Spanish speakers*[18]. Dampak interaksi dengan LLM terhadap persepsi diri juga diteliti menggunakan transkrip interaksi antara partisipan dan LLM, survei persepsi diri sebelum dan sesudah interaksi, serta data kualitatif dari wawancara mendalam[20].

Evaluasi komparatif model GPT dilakukan menggunakan sampel representatif dari UK dengan 401 peserta yang disaring menjadi 365 partisipan akhir dengan rata-rata usia 46,9 tahun. Penelitian ini membandingkan respons dari tiga model GPT: GPT-3.5-turbo-0125, GPT-4-0612, dan GPT-4-0125-preview[1]. Dataset pelatihan dan evaluasi yang komprehensif juga digunakan, mencakup data teks dari domain umum dan spesifik untuk menguji kemampuan pemahaman konteks, jawaban atas soal, dan penyelesaian tugas berbasis bahasa menggunakan korpus besar yang umum digunakan dalam pelatihan LLM[21].

Pendekatan filosofis dalam memahami LLM dilakukan melalui analisis teks dan konsep filosofis dari karya Heidegger dan Lacan, dengan karakterisasi dan analisis fungsi LLM seperti *Generative Pre-trained Transformer* (GPT). Penelitian ini menggunakan observasi fenomenologis tentang interaksi manusia dengan LLM serta contoh-contoh percakapan dan output yang menunjukkan sifat dan keterbatasan model[3].

Implementasi LLM dalam konteks industri menggunakan kumpulan data yang beragam termasuk laporan perusahaan, survei tenaga kerja, deskripsi pekerjaan, dan data publik lainnya yang relevan dengan praktik pemberi kerja[24]. Dataset lain mencakup kumpulan teks besar dari berbagai domain publik dan spesifik industri untuk melatih dan menguji performa LLM[22], serta kumpulan laporan kasus, teks naratif, dan data sintetis untuk menguji kemampuan LLM dalam memahami dan menghasilkan laporan berbasis teks[23]. Evaluasi kemampuan kognitif juga dilakukan menggunakan tugas-tugas spesifik dalam domain logika, matematika, bahasa, dan pengetahuan umum untuk pengujian model dalam berbagai konteks[13].

3.5. Hasil Penelitian

Large language models mendemonstrasikan kapabilitas yang signifikan dalam menganalisis konten tekstual dengan akurasi tinggi. Dalam konteks analisis naratif pasien, LLM berhasil mengekstrak tema-tema penting dengan persentase identifikasi yang mengesankan, dimana 95% naratif mengandung tema perawatan, 67% terkait komunikasi, dan 51% membahas lingkungan rumah sakit. Model-model ini juga menunjukkan kemampuan superior dalam mendeteksi nada emosional dengan distribusi 80% positif, 7% negatif, dan 13% campuran, serta mampu mengidentifikasi perasaan dominan seperti rasa terima kasih (42%) dan kepuasan (30%)[2].

Keunggulan LLM dalam pemrosesan bahasa natural juga tercermin pada kemampuannya mengidentifikasi pola-pola kompleks dalam data tekstual, termasuk kebutuhan keterampilan yang berkembang, tren perekrutan, dan preferensi pemberi kerja[24]. Model-model ini dapat menganalisis *subspace* semantik yang efektif dengan dimensi terbatas (20-50) untuk menangkap perbedaan linguistik terkait kepribadian, mencapai akurasi yang sebanding dengan metode tradisional berbasis fitur leksikal namun dengan dengan kemampuan interpretabilitas yang superior[18].

Penelitian menunjukkan bahwa LLM memiliki karakteristik kecerdasan yang menyerupai manusia, baik secara individual maupun kolektif. GPT-4 mengungguli performa rata-rata manusia dengan akurasi 83% dibandingkan 64% untuk manusia dalam tugas-tugas kognitif individual. Lebih mengesankan lagi, ketika jawaban dari multiple LLM diagregasi, mereka mencapai tingkat akurasi kolektif hingga 97.5% untuk GPT-4, jauh melampaui kecerdasan kolektif manusia yang hanya mencapai 86%. Teknik agregasi seperti pembobotan berdasarkan *confidence level* terbukti dapat meningkatkan performa kolektif LLM secara signifikan[15].

Model-model ini juga menunjukkan kemampuan kognitif yang saling terkait, mencakup penalaran logis, pemahaman konteks, dan kemampuan generalisasi yang muncul dari proses training[13]. LLM mampu mereplikasi aspek-aspek tertentu dari penalaran induktif manusia, terutama dalam tugas-tugas yang melibatkan pola umum dan hubungan semantik[14].

Dalam domain klinis, LLM menunjukkan potensi yang menjanjikan untuk mendukung *assessment* dan diagnosis psikologis. Model-model ini mampu melakukan stratifikasi risiko bunuh diri dengan performa yang sebanding atau superior dibandingkan model machine *learning tradisional*, dengan GPT-4 mencapai akurasi

p-ISSN: 2775-4227
e-ISSN: 2775-4219

76.2% dan korelasi Spearman 0.816 dengan penilaian klinis. Semua LLM menunjukkan sensitivitas tinggi (>90%) dalam mengidentifikasi risiko bunuh diri level tinggi[9].

LLM juga mendemonstrasikan kemampuan mengidentifikasi dan menilai dimensi gangguan pemikiran dengan tingkat akurasi yang menjanjikan. GPT-4 menunjukkan performa superior dibandingkan Claude dalam mengidentifikasi gangguan pemikiran formal, dengan korelasi signifikan terhadap penilaian expert manusia (r=0.70). Model-model ini berhasil membedakan antara pasien dengan gangguan skizofrenia dan bipolar dari kontrol sehat, serta dapat mengidentifikasi subtipe gangguan pemikiran spesifik yang sesuai dengan diagnosis klinis[10].

Untuk assessment OCD, ChatGPT menunjukkan kemampuan dalam menghasilkan hierarki dengan performa yang memadai, meskipun masih inferior dibandingkan tenaga ahli manusia. Model ini mencapai rating *appropriateness* (M = 4.47), *specificity* (M = 4.17), dan *safety/ethics* (M = 4.89) yang cukup tinggi, namun *human-generated hierarchies* tetap mendapat rating yang signifikan lebih tinggi dalam hampir semua aspek[12].

Kemampuan LLM dalam mendeteksi nuansa bahasa menunjukkan hasil yang bervariasi. ChatGPT memiliki kapabilitas yang memadai dalam mengenali ketidaksopanan dalam teks, dengan akurasi yang lebih tinggi untuk ketidaksopanan eksplisit dibandingkan yang bersifat implisit atau kontekstual. Model komersial umumnya dapat mengenali dan menolak prasangka eksplisit, namun sering gagal mengidentifikasi dan bahkan mendukung prasangka implisit, sementara model *open-source* cenderung lebih sering menerima prasangka baik implisit maupun eksplisit[8].

Meskipun menunjukkan kemampuan yang mengesankan, LLM menghadapi beberapa keterbatasan signifikan. Model-model ini sering menghasilkan respons yang tampak masuk akal tetapi sebenarnya tidak memiliki dasar logis atau pemahaman yang benar, fenomena yang dikenal sebagai "*cognitive phantoms*". LLM lebih mengandalkan pola statistik daripada pemahaman semantik yang sesungguhnya[1].

Dalam konteks *moral reasoning*, LLM memiliki kemampuan terbatas dalam memahami dan menerapkan nilai-nilai moral secara konsisten. Meskipun dapat memberikan respons yang tampak etis dalam beberapa kasus, mereka sering gagal menangkap nuansa moral yang kompleks atau memberikan jawaban yang sesuai dengan konteks spesifik[16]. Model-model ini juga dapat menunjukkan bias signifikan dalam responsnya, yang encerminkan bias dalam data training atau pola interaksi pengguna[17].

Keterbatasan lain meliputi kesulitan dalam memahami konteks yang sangat kompleks atau spesifik[11], [22], [23], kebutuhan komputasi yang tinggi[22], dan tantangan dalam menangani tugas yang memerlukan pemahaman mendalam tentang konteks atau pengetahuan dunia nyata[13], [14]. Penelitian juga menunjukkan bahwa peningkatan ukuran model tidak selalu memberikan keuntungan yang signifikan, dan efisiensi serta tujuan spesifik harus menjadi pertimbangan utama[21].

Interaksi dengan LLM dapat memiliki efek beragam pada persepsi diri individu. Beberapa partisipan melaporkan peningkatan rasa percaya diri dan pemahaman diri setelah berinteraksi dengan LLM, sementara yang lain merasa kurang percaya diri atau mengalami kebingungan terkait identitas mereka[20]. Temuan ini menyoroti pentingnya desain etis dan bertanggung jawab dalam pengembangan teknologi berbasis LLM.

4. KESIMPULAN

Tinjauan literatur sistematis ini memberikan gambaran komprehensif mengenai penerapan *large language models* (LLM) dalam bidang psikologi. Melalui analisis dan sintesis temuan dari 20 studi terkini, beberapa kesimpulan penting dapat ditarik.

Pertama, LLM menunjukkan kemampuan yang signifikan dalam menganalisis dan memproses konten tekstual dengan akurasi tinggi, khususnya dalam konteks analisis naratif pasien dan identifikasi pola emosional. Model-model ini juga mendemonstrasikan karakteristik kecerdasan individual dan kolektif yang dalam beberapa aspek menyamai atau bahkan melampaui performa manusia, terutama ketika diimplementasikan secara kolektif melalui teknik agregasi.

Kedua, dalam aplikasi klinis, LLM menunjukkan potensi yang menjanjikan untuk mendukung *assessment* dan diagnosis psikologis, termasuk stratifikasi risiko bunuh diri dan identifikasi gangguan pemikiran. Namun, performa mereka masih belum konsisten mencapai standar tenaga ahli manusia dalam semua aspek *assessment* klinis.

Ketiga, meskipun memiliki kemampuan yang mengesankan, LLM menghadapi keterbatasan fundamental dalam hal pemahaman semantik yang sesungguhnya, penalaran moral, dan bias yang dapat memengaruhi reliabilitas output mereka. Fenomena "*cognitive phantoms*" dan ketergantungan pada pola statistik daripada pemahaman kontekstual yang mendalam menjadi perhatian utama.

Keempat, dampak psikososial dari interaksi dengan LLM bervariasi antar individu, dengan potensi efek positif dan negatif yang perlu dipertimbangkan dalam implementasi teknologi ini dalam dunia psikologis.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Berdasarkan temuan-temuan tersebut, dapat disimpulkan bahwa LLM memiliki potensi besar untuk mendukung praktik psikologi, namun implementasinya harus dilakukan dengan hati-hati, selalu disertai pengawasan manusia yang kompeten, dan dengan pertimbangan etis yang matang. Penelitian lebih lanjut diperlukan untuk mengatasi keterbatasan yang ada dan mengoptimalkan manfaat teknologi ini dalam layanan psikologis.

Temuan penelitian ini memberikan kontribusi teoretis dengan memperkaya pemahaman tentang kemampuan komputasional dalam memproses fenomena psikologis dan membuka diskusi baru mengenai batas-batas AI dalam memahami kompleksitas mental manusia. Secara praktis, hasil ini memberikan panduan konkret bagi praktisi psikologi untuk mengadopsi LLM sebagai alat bantu assessment dan diagnosis melalui pendekatan hibrid yang menggabungkan kekuatan AI dengan keahlian klinis manusia. Implikasi praktis mencakup pengembangan protokol standar implementasi LLM dalam setting klinis, pelatihan profesional dalam penggunaan AI, dan perumusan guidelines etis yang komprehensif. Penelitian ini juga mengidentifikasi area prioritas pengembangan teknologi masa depan, termasuk peningkatan pemahaman kontekstual dan pengurangan bias algoritmik untuk menginformasikan agenda riset psikologi komputasional selanjutnya.

DAFTAR PUSTAKA

[1] S. Peereboom, I. Schwabe, and B. Kleinberg, "Cognitive phantoms in large language models through the lens of latent variables," *Computers in Human Behavior: Artificial Humans*, vol. 4, p. 100161, May 2025, doi: 10.1016/j.chbah.2025.100161.

[2] S. Jenner *et al.*, "Using large language models for narrative analysis: a novel application of generative AI," *Methods in Psychology*, vol. 12, p. 100183, Jun. 2025, doi: 10.1016/j.metip.2025.100183.

[3] M. Heimann and A.-F. Hübener, "Circling the void: Using Heidegger and Lacan to think about large language models," *Cognitive Systems Research*, vol. 91, p. 101349, Jun. 2025, doi: 10.1016/j.cogsys.2025.101349.

[4] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Journal of Clinical Epidemiology*, vol. 134, pp. 178–189, Jun. 2021, doi: 10.1016/j.jclinepi.2021.03.001.

[5] Z. Guo, A. Lai, J. H. Thygesen, J. Farrington, T. Keen, and K. Li, "Large Language Models for Mental Health Applications: Systematic Review," *JMIR Ment Health*, vol. 11, p. e57400, Oct. 2024, doi: 10.2196/57400.

[6] A. Ferrario, J. Sedlakova, and M. Trachsel, "The Role of Humanization and Robustness of Large Language Models in Conversational Artificial Intelligence for Individuals With Depression: A Critical Analysis," *JMIR Ment Health*, vol. 11, pp. e56569–e56569, Jul. 2024, doi: 10.2196/56569.

[7] O. N. E. Kjell, K. Kjell, and H. A. Schwartz, "Beyond rating scales: With targeted evaluation, large language models are poised for psychological assessment," *Psychiatry Research*, vol. 333, p. 115667, Mar. 2024, doi: 10.1016/j.psychres.2023.115667.

[8] T. H. McCoy and R. H. Perlis, "Characterizing research domain criteria symptoms among psychiatric inpatients using large language models," *Journal of Mood & Anxiety Disorders*, vol. 8, p. 100079, Dec. 2024, doi: 10.1016/j.xjmad.2024.100079.

[9] T. H. McCoy and R. H. Perlis, "Applying large language models to stratify suicide risk using narrative clinical notes," *Journal of Mood & Anxiety Disorders*, vol. 10, p. 100109, Jun. 2025, doi: 10.1016/j.xjmad.2025.100109.

[10] S. L. Pugh, C. Chandler, A. S. Cohen, C. Diaz-Asper, B. Elvevåg, and P. W. Foltz, "Assessing dimensions of thought disorder with large language models: The tradeoff of accuracy and consistency," *Psychiatry Research*, vol. 341, p. 116119, Nov. 2024, doi: 10.1016/j.psychres.2024.116119.

[11] L. M. Vowels, S. Sweeney, and M. J. Vowels, "Evaluating the Efficacy of Amanda: A Voice-Based Large Language Model Chatbot for Relationship Challenges," Dec. 22, 2024. doi: 10.31234/osf.io/3x7e8.

[12] E. E. Bernstein *et al.*, "Feasibility of Using ChatGPT to Generate Exposure Hierarchies for Treating Obsessive-Compulsive Disorder," *Behavior Therapy*, p. S0005789425000231, Mar. 2025, doi: 10.1016/j.beth.2025.02.005.

[13] D. Ilić and G. E. Gignac, "Evidence of interrelated cognitive-like capabilities in large language models: Indications of artificial general intelligence or achievement?," *Intelligence*, vol. 106, p. 101858, Sep. 2024, doi: 10.1016/j.intell.2024.101858.

[14] S. J. Han, K. J. Ransom, A. Perfors, and C. Kemp, "Inductive reasoning in humans and large language models," *Cognitive Systems Research*, vol. 83, p. 101155, Jan. 2024, doi: 10.1016/j.cogsys.2023.101155.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[15] L. Sun *et al.*, "Large language models show both individual and collective creativity comparable to humans," *Thinking Skills and Creativity*, vol. 57, p. 101870, Sep. 2025, doi: 10.1016/j.tsc.2025.101870.

[16] L. Bulla, S. De Giorgis, M. Mongiovì, and A. Gangemi, "Large Language Models meet moral values: A comprehensive assessment of moral abilities," *Computers in Human Behavior Reports*, vol. 17, p. 100609, Mar. 2025, doi: 10.1016/j.chbr.2025.100609.

[17] Z. W. Petzel and L. Sowerby, "Prejudiced interactions with large language models (LLMs) reduce trustworthiness and behavioral intentions among members of stigmatized groups," *Computers in Human Behavior*, vol. 165, p. 108563, Apr. 2025, doi: 10.1016/j.chb.2025.108563.

[18] J. Á. Martínez-Huertas, G. Jorge-Botana, A. Martínez-Mingo, J. D. Moreno, and R. Olmos, "Modeling personality language use with small semantic vector subspaces," *Personality and Individual Differences*, vol. 219, p. 112514, Mar. 2024, doi: 10.1016/j.paid.2023.112514.

[19] M. Andersson and D. McIntyre, "Can ChatGPT recognize impoliteness? An exploratory study of the pragmatic awareness of a large language model," *Journal of Pragmatics*, vol. 239, pp. 16–36, Apr. 2025, doi: 10.1016/j.pragma.2025.02.001.

[20] O. L. Jacobs, F. Pazhoohi, and A. Kingstone, "Large language models have divergent effects on self-perceptions of mind and the attributes considered uniquely human," *Consciousness and Cognition*, vol. 124, p. 103733, Sep. 2024, doi: 10.1016/j.concog.2024.103733.

[21] E. G. Wilcox *et al.*, "Bigger is not always better: The importance of human-scale language modeling for psycholinguistics," Jul. 17, 2024. doi: 10.31234/osf.io/rfwgd.

[22] A. Bewersdorff *et al.*, "Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education," *Learning and Individual Differences*, vol. 118, p. 102601, Feb. 2025, doi: 10.1016/j.lindif.2024.102601.

[23] D. Stoll, S. Wehrli, and D. Lätsch, "Case reports unlocked: Harnessing large language models to advance research on child maltreatment," *Child Abuse & Neglect*, vol. 160, p. 107202, Feb. 2025, doi: 10.1016/j.chiabu.2024.107202.

[24] A. Grybauskas and J. Cárdenas-Rubio, "Unlocking employer insights: Using large language models to explore human-centric aspects in the context of industry 5.0," *Technological Forecasting and Social Change*, vol. 208, p. 123719, Nov. 2024, doi: 10.1016/j.techfore.2024.123719.