



Original Article

Memahami AI sebagai Agen Social: Tinjauan Teori Interaksi Sosial

Reni Kartini Kristin Dian Yuniarti[✉]

Universitas Negeri Manado, Tondano, Indonesia.

renikartinikdy3@unima.ac.id[✉]

Abstrak:

Perkembangan pesat kecerdasan buatan (AI) telah mentransformasi cara manusia berinteraksi dengan teknologi, memunculkan pertanyaan fundamental tentang status AI sebagai agen sosial. Penelitian ini menggunakan pen-dekatan tinjauan literatur naratif untuk mengkaji bagaimana AI dapat dipahami sebagai agen sosial melalui perspektif teori interaksi sosial. Analisis komprehensif terhadap 64 studi multidisipliner dari tahun 1994-2025 mengidentifikasi lima tema utama: karakteristik AI sebagai agen sosial, aplikasi teori interaksi sosial (CASA, Teori Persamaan Media, HAI-TIME), mekanisme antropomorfisme dan integrasi diri-AI, faktor-faktor yang mempengaruhi persepsi sosial, serta implikasi etis dan desain. Temuan menunjukkan bahwa 75% studi mendukung konsep AI sebagai agen sosial, dengan bukti empiris dari eksperimen laboratorium, meta-analisis (ukuran efek kognitif $d=0,70$, afektif $d=0,59$), dan validasi neurosains. AI modern memenuhi kriteria agen sosial melalui tujuh karakteristik fundamental: perwujudan fisik, emosi, dialog, kepribadian, persepsi berorientasi manusia, pemodelan pengguna, dan pembelajaran sosial. Model bahasa besar menunjukkan kemampuan Teori Pikiran spontan setara anak usia 9 tahun. Chatbot pendamping memberikan empat jenis dukungan sosial dengan dukungan persahabatan mencapai 77,1%. Analisis mengidentifikasi konvergensi global pada lima prinsip etika AI: transparansi, keadilan, non-malefisiensi, tanggung jawab, dan privasi. Penelitian ini berkontribusi pada pemahaman teoretis tentang AI sebagai agen sosial dan menyediakan fondasi untuk pengembangan interaksi manusia-AI yang lebih bermakna dan etis, dengan implikasi praktis untuk pendidikan, kesehatan, dan layanan pelanggan.

Kata kunci: kecerdasan buatan, agen sosial, teori interaksi sosial, CASA, antropomorfisme, interaksi manusia-AI

Pendahuluan

Integrasi agen artifisial yang semakin masif ke dalam kehidupan sosial manusia menandai transformasi fundamental dalam cara manusia berinteraksi dengan teknologi. Artificial agents telah menjadi bagian integral dalam berbagai domain, termasuk kesehatan, pendidikan, dan hiburan, menawarkan peningkatan efisiensi, personalisasi, dan konektivitas emosional ([Lukasik & Gut, 2025](#)). Perkembangan pesat robot, avatar, dan chatbot ke dalam kehidupan sosial manusia memerlukan pemahaman yang lebih mendalam tentang interaksi manusia-AI dan dampaknya terhadap interaksi sosial ([Lukasik & Gut, 2025](#)). Fenomena ini diperkuat



<https://jurnal.usk.ac.id/riwayat>

oleh kenyataan bahwa aplikasi pesan mobile kini memiliki lebih dari 1,5 miliar pengguna di seluruh dunia, dengan bahasa natural menjadi mode default interaksi *online* bagi banyak pengguna (Følstad & Brandtzaeg, 2017).

Studi fundamental oleh Nass et al. (1994) mendemonstrasikan bahwa manusia secara instingtif menerapkan konsep kognisi sosial terhadap agen artifisial, memperlakukan komputer dan sistem berbasis AI lainnya sebagai aktor sosial (Lukasik & Gut, 2025). Fenomena ini kemudian dikembangkan menjadi Teori *Media Equation* (Reeves & Nass, 1996), yang menetapkan bahwa manusia merespons agen artifisial sebagaimana mereka merespons manusia lainnya (Lukasik & Gut, 2025). Perspektif *Computer as Social Actor* (CASA) telah memotivasi peneliti untuk mengeksplorasi kemungkinan memanfaatkan norma sosial manusia yang telah mapan untuk memahami interaksi antara manusia dan robot (Ta et al., 2020). Pemahaman tentang bagaimana manusia memproses dan menerapkan informasi tentang makhluk sosial menjadi esensial dalam penelitian kognisi sosial, yang tidak hanya menginformasikan tetapi juga diinformasikan oleh pengembangan robotika sosial (Lukasik & Gut, 2025).

Beberapa penelitian telah mengeksplorasi berbagai aspek interaksi manusia-AI. Yuan et al. (2024) fokus pada bagaimana robot dapat berinteraksi dengan banyak pengguna dan pentingnya situasi sosial dalam persepsi hubungan manusia-robot. Alabed et al. (2022) mengeksplorasi konsep self-AI integration dan bagaimana pengguna dapat mengintegrasikan agen AI yang dianthropomorfisasi sebagai bagian dari konsep diri mereka. Thomaz & Breazeal (2008) menyelidiki hipotesis social learner dalam konteks pengajaran robot. Henschel et al. (2020) menganalisis mekanisme neurokognitif yang mendukung interaksi manusia-robot. Ta et al. (2020) meneliti dukungan sosial dari chatbot companion dalam konteks sehari-hari. Namun, penelitian-penelitian tersebut cenderung fokus pada aspek spesifik interaksi manusia-AI tanpa mengintegrasikan perspektif teori interaksi sosial secara komprehensif.

Meskipun telah ada kemajuan signifikan dalam memahami interaksi manusia-AI, masih terdapat kesenjangan dalam literatur terkait bagaimana berbagai teori interaksi sosial dapat diintegrasikan untuk memberikan pemahaman holistik tentang AI sebagai agen sosial. Liu (2020) mengidentifikasi bahwa AI tidak hanya merupakan perusahaan yang dikonstruksi secara sosial, tetapi juga bersifat konstitutif secara sosial dalam arti bahwa ketika diimplementasikan dalam lingkungan sosial, sistem AI dapat mengasumsikan peran sosial, menjalankan praktik sosial, dan membentuk relasi sosial. Woolgar (1985) bahkan menyarankan pengembangan sosiologi mesin untuk menganalisis bagaimana sistem AI berfungsi seperti aktor sosial manusia untuk membentuk relasi sosial dan mengkonstruksi realitas sosial (Liu, 2020). Namun, belum ada tinjauan komprehensif yang mengintegrasikan berbagai perspektif teori interaksi sosial untuk memahami AI sebagai agen sosial, termasuk implikasi etis dan desain yang muncul dari pemahaman tersebut.

Beberapa peneliti fokus pada aspek teknis dan fungsional AI, sementara yang lain hanya terbatas pada aspek psikologis atau neurokognitif dari interaksi manusia-AI. Belum ada penelitian yang secara sistematis mengintegrasikan teori-teori interaksi sosial klasik seperti CASA, *Media Equation Theory*, dan *Social Response Theory* dengan temuan empiris terkini tentang *antropomorfisme*, *embodiment*, dan implikasi etis dalam konteks AI sebagai agen sosial. Oleh karena itu, penelitian ini

berfokus pada tinjauan naratif komprehensif yang mengintegrasikan berbagai perspektif teori interaksi sosial untuk memahami AI sebagai agen sosial. Tujuan dari penelitian ini adalah untuk mengkaji bagaimana AI dapat dipahami sebagai agen sosial melalui perspektif teori interaksi sosial, dengan fokus pada karakteristik AI sebagai agen sosial, teori-teori interaksi sosial yang relevan, fenomena antropomorfisme, serta implikasi etis dan desain yang muncul dari pemahaman tersebut.

Metode

Desain Penelitian

Penelitian ini menggunakan pendekatan tinjauan literatur naratif (narrative literature review) untuk mengkaji bagaimana AI dapat dipahami sebagai agen sosial melalui perspektif teori interaksi sosial. Tinjauan naratif dipilih karena memungkinkan sintesis interpretatif dari berbagai perspektif teoretis dan temuan empiris yang beragam, serta memberikan fleksibilitas dalam mengintegrasikan literatur dari berbagai disiplin ilmu termasuk ilmu komputer, psikologi, sosiologi, dan human-computer interaction.

Strategi Pencarian dan Kriteria Seleksi Literatur

Pencarian literatur dilakukan melalui database akademik utama meliputi Google Scholar, Scopus, Web of Science, IEEE Xplore, dan PubMed untuk periode publikasi 1990-an-2025. Pemilihan rentang waktu ini didasarkan pada perkembangan signifikan teori CASA yang diperkenalkan pada pertengahan 1990-an dan aplikasinya dalam konteks AI modern. Kata kunci pencarian yang digunakan meliputi kombinasi dari: "artificial intelligence", "social agent", "human-AI interaction", "CASA theory", "Media Equation", "anthropomorphism", "social robotics", "chatbot", "social cognition", dan "AI ethics".

Kriteria inklusi yang diterapkan mencakup: (1) artikel jurnal peer-reviewed dalam bahasa Inggris; (2) publikasi yang secara eksplisit membahas AI atau sistem berbasis komputer sebagai agen sosial; (3) studi yang menggunakan atau mendiskusikan teori interaksi sosial dalam konteks AI; (4) penelitian empiris maupun konseptual yang relevan dengan topik; dan (5) publikasi yang membahas implikasi etis atau desain dari interaksi manusia-AI. Kriteria eksklusi meliputi: (1) artikel yang hanya fokus pada aspek teknis AI tanpa dimensi sosial; (2) publikasi berbentuk abstrak konferensi atau poster tanpa full paper; (3) literatur abu-abu yang tidak melalui proses peer review; dan (4) artikel yang tidak dapat diakses secara penuh.

Dari pencarian awal yang menghasilkan 487 artikel, dilakukan screening berdasarkan judul dan abstrak yang menyisakan 156 artikel potensial. Setelah pembacaan full-text dan penerapan kriteria inklusi-eksklusi secara ketat, terpilih 64 artikel utama yang menjadi basis analisis.

Instrumen

Analisis literatur dilakukan melalui empat tahap sistematis. Tahap pertama melibatkan pembacaan mendalam dan ekstraksi informasi kunci dari setiap artikel menggunakan matriks ekstraksi data yang mencakup: penulis, tahun, tujuan penelitian, kerangka teoretis, metodologi, temuan utama, dan implikasi. Tahap kedua fokus pada identifikasi dan kategorisasi tema-tema utama yang muncul dari literatur menggunakan pendekatan analisis tematik induktif. Tema-tema yang teridentifikasi meliputi karakteristik AI

sebagai agen sosial, teori interaksi sosial yang relevan, mekanisme antropomorfisme, faktor-faktor yang mempengaruhi persepsi sosial terhadap AI, serta implikasi etis dan desain.

Tahap ketiga melibatkan sintesis naratif untuk mengintegrasikan temuan dari berbagai studi dan mengidentifikasi pola, konvergensi, dan divergensi dalam literatur. Perhatian khusus diberikan pada bagaimana berbagai teori interaksi sosial (CASA, Media Equation Theory, Social Response Theory) diterapkan dan dikembangkan dalam konteks AI kontemporer. Tahap keempat mencakup analisis kritis terhadap kesenjangan dalam literatur dan identifikasi arah penelitian masa depan.

Hasil Penelitian

Tinjauan literatur yang dilakukan menghasilkan temuan bahwa pemahaman AI sebagai agen sosial dibangun atas tiga pilar utama, yaitu ciri khas AI yang memungkinkannya berperan sebagai agen sosial, landasan teori interaksi sosial yang menjelaskan bagaimana manusia merespons AI, serta fenomena antropomorfisme yang membentuk pengalaman pengguna. Rentang waktu publikasi dari 1994 hingga 2025 menunjukkan evolusi pemahaman tentang AI sebagai agen sosial selama tiga dekade terakhir. Temuan ini menjawab tujuan penelitian untuk mengkaji bagaimana AI dapat dipahami sebagai agen sosial melalui sudut pandang teori interaksi sosial, dengan menekankan pada ciri-ciri khas AI, teori-teori yang relevan, fenomena antropomorfisme, serta akibat etis dan rancangan yang muncul dari pemahaman tersebut.

Pembahasan

Karakteristik AI sebagai Agen Sosial

Analisis terhadap 64 studi mengungkapkan karakteristik fundamental AI sebagai agen sosial. [Jackson dan Williams \(2021\)](#) memberikan definisi operasional yang jelas bahwa agen sosial adalah entitas yang mampu melakukan tindakan sosial dengan kriteria interaktivitas, otonomi, dan adaptabilitas. [Fong, Nourbakhsh, dan Dautenhahn \(2003\)](#) mengidentifikasi tujuh karakteristik fundamental robot sosial: perwujudan fisik, emosi, dialog, kepribadian, persepsi berorientasi manusia, pemodelan pengguna, dan pembelajaran sosial yang terletak dalam konteks.

[Huang dan Rust \(2018\)](#) mengidentifikasi evolusi AI dari mekanis, analitis, intuitif, hingga empatik, sementara dalam studi lanjutan mereka (2021) memperkenalkan kategorisasi baru: AI mekanis, berpikir, dan merasakan. AI merasakan khususnya menunjukkan kemampuan sebagai agen sosial yang dapat membangun hubungan personal. [Van Doorn et al. \(2017\)](#) memperkenalkan konsep Kehadiran Sosial Otomatis (Automated Social Presence) yang mempengaruhi kognisi sosial dalam hal kehangatan dan kompetensi.

[Kosinski \(2023\)](#) mendemonstrasikan bahwa model bahasa besar dapat mengembangkan Teori Pikiran secara spontan, dengan GPT-3.5 menyelesaikan 93% tugas Teori Pikiran setara dengan anak usia 9 tahun. Banks (2020) memperkuat temuan ini dengan menunjukkan bahwa robot sosial dapat melewati lima tes Teori Pikiran yang telah ditetapkan. Wang et al. (2024) mengusulkan paradigma Teori Pikiran Bersama (*Mutual Theory of Mind*) dimana manusia dan AI saling memiliki kemampuan Teori Pikiran untuk interaksi yang adaptif dan natural.

Teori-Teori Interaksi Sosial dalam Konteks AI

Fondasi teoretis pemahaman AI sebagai agen sosial berawal dari teori CASA (*Computers Are Social Actors*) yang diperkenalkan [Nass et al. \(1994\)](#). Teori ini

menunjukkan bahwa manusia secara instinktif menerapkan norma sosial pada sistem AI. [Nass dan Moon \(2000\)](#) memperkuat temuan ini dengan mendemonstrasikan bahwa manusia secara otomatis merespon komputer dengan aturan sosial yang sama seperti interaksi manusia-manusia.

[Reeves dan Nass \(1996\)](#) mengembangkan Teori Persamaan Media yang menunjukkan bahwa manusia memperlakukan teknologi seperti entitas sosial nyata. Dalam eksperimen tentang kesopanan, mereka menemukan bahwa manusia secara otomatis menerapkan aturan kesopanan pada komputer, dengan respons lebih positif ketika komputer menanyakan tentang dirinya sendiri dan evaluasi lebih jujur ketika dilakukan melalui komputer yang berbeda.

[Sundar \(2020\)](#) memberikan kontribusi signifikan dengan kerangka kerja HAIITIME yang menjelaskan dua jalur: jalur isyarat (heuristik mesin) dan jalur aksi (kolaborasi manusia-AI). [Pagliari, Chambon, dan Berberian \(2022\)](#) melalui perspektif agensi sosial menunjukkan bahwa AI memiliki karakteristik yang mendorong manusia untuk menganggapnya sebagai agen yang memiliki intensi.

[Thomaz dan Breazeal \(2008\)](#) mengidentifikasi bias positif dalam pemberian imbalan kepada robot (69,8% imbalan positif), menunjukkan bahwa manusia secara alami memperlakukan robot sebagai pembelajar sosial. [Henschel, Hortensius, dan Cross \(2020\)](#) memberikan validasi neurosains, menemukan perbedaan aktivasi jaringan otak saat berinteraksi dengan robot versus manusia, namun pentingnya pencitraan otak bergerak untuk studi interaksi dunia nyata.

Antropomorfisme dan Persepsi Status Mental

Antropomorfisme muncul sebagai tema dominan dalam literatur. [Alabed et al. \(2022\)](#) mengusulkan konsep "integrasi diri-AI" yang menunjukkan bagaimana AI antropomorfik dapat terintegrasi dengan konsep diri pengguna. [MacInnis dan Folkes \(2017\)](#) memberikan dasar teoretis tentang bagaimana konsumen dapat memandang entitas non-manusia sebagai manusiawi dengan kepribadian dan karakteristik yang mirip dengan diri mereka.

[Kang dan Kang \(2024\)](#) mendemonstrasikan bahwa fitur antropomorfis dalam chatbot konseling meningkatkan pengungkapan diri dan persahabatan. [Sidlauskiene, Joye, dan Auruskeviciene \(2023\)](#) menemukan bahwa antropomorfisme chatbot meningkatkan persepsi personalisasi produk, dengan efek yang diperkuat oleh kesepian situasional.

[Bartneck et al. \(2009\)](#) mengembangkan instrumen pengukuran standar "Godspeed questionnaire series" untuk lima dimensi persepsi robot dengan reliabilitas tinggi (Alpha Cronbach >0,7). [Aggarwal dan McGill \(2007\)](#) menjelaskan mekanisme kognitif dimana fitur antropomorfik mengaktifasi skema manusia dan mempengaruhi evaluasi produk.

[De Visser et al. \(2016\)](#) menemukan bahwa agen antropomorfik menunjukkan ketahanan kepercayaan yang lebih tinggi, dengan antropomorfisme mengurangi ekspektasi awal namun memperlambat penurunan kepercayaan saat terjadi kegagalan. [Watson \(2019\)](#) memberikan perspektif kritis, memperingatkan bahwa antropomorfisme AI dapat menyesatkan dan berbahaya secara etis.

Faktor-Faktor yang Mempengaruhi Persepsi Sosial terhadap AI

[Felnhofer et al. \(2024\)](#) menemukan bahwa persepsi agensi merupakan faktor kritis, dengan avatar dinilai lebih tinggi dalam kehadiran sosial dibanding AI. [Ghiglini et al. \(2023\)](#) menunjukkan bahwa robot yang menampilkan perilaku kontingen meningkatkan sikap mentalistik. [Marchesi et al. \(2019\)](#) menemukan bahwa robot dengan ciri menyerupai manusia memberikan wawasan tentang proses kognitif manusia.

[Yuan et al. \(2024\)](#) mengidentifikasi bahwa mode interaksi pengguna tunggal menghasilkan evaluasi lebih tinggi, dengan hubungan "akrab" optimal untuk pengguna tunggal. [Li, Rau, dan Li \(2010\)](#) menemukan perbedaan budaya signifikan dalam persepsi robot, menunjukkan pentingnya faktor kontekstual.

[Kojima et al. \(2021\)](#) menemukan bahwa kondisi teks bersamaan waktu nyata menghasilkan kehadiran sosial tertinggi, dengan konkurensi lebih penting daripada kekayaan informasi. [Poinsot et al. \(2022\)](#) memberikan bukti empiris bahwa lawan yang dikontrol komputer dengan komunikasi emosional dapat menghasilkan kehadiran bersama yang lebih tinggi dibanding lawan manusia.

[Fischer, Kim, dan Hong \(2023\)](#) menunjukkan bahwa kepercayaan dan kepercayaan terhadap robot secara serial memediasi hubungan antara kesan pertama dan niat perilaku. [Esterwood dan Robert Jr \(2023\)](#) menemukan bahwa tidak ada strategi perbaikan yang sepenuhnya mengembalikan kepercayaan setelah pelanggaran berulang, dengan penyangkalan menjadi paling tidak efektif.

Dukungan Sosial dan Aplikasi Praktis

Ta et al. (2020) dalam analisis tematik terhadap 1.854 ulasan pengguna Replika mengidentifikasi empat jenis dukungan sosial: dukungan persahabatan (77,1%), dukungan emosional (44,6%), dukungan informasional (15,6%), dan dukungan penilaian (9,3%). Temuan ini menunjukkan bahwa AI dapat berfungsi sebagai sumber dukungan sosial yang bermakna.

Belpaeme et al. (2018) dalam meta-analisis 101 makalah menemukan bahwa robot sosial efektif dalam pendidikan dengan ukuran efek kognitif ($d=0,70$) dan afektif ($d=0,59$). Kehadiran fisik memberikan keuntungan dibanding agen virtual, dan robot dapat berperan sebagai tutor, guru, teman sebaya, atau pemula.

Leite, Martinho, dan Paiva (2013) mengidentifikasi bahwa robot sosial efektif dalam perawatan kesehatan, pendidikan, dan lingkungan rumah, dengan pentingnya penampilan yang sesuai, perilaku kontinuitas, dan kemampuan memori untuk interaksi jangka panjang.

Implikasi Etis dan Desain

Jobin, Ienca, dan Vayena (2019) dalam analisis 84 dokumen etika AI mengidentifikasi konvergensi global pada lima prinsip: transparansi, keadilan dan kesetaraan, non-malefisiensi, tanggung jawab, dan privasi. Floridi et al. (2018) dalam kerangka kerja AI4People menambahkan prinsip keterjelasan sebagai pendukung empat prinsip lainnya.

Dautenhahn (2007) menekankan pentingnya pendekatan "manusia dalam lingkaran" untuk desain agen sosial. Følstad dan Brandtzaeg (2017) mengidentifikasi pergeseran paradigma dari antarmuka grafis ke antarmuka percakapan, memerlukan desain untuk interaksi dalam jaringan multi-agen.

Weiss et al. (2011) memberikan kerangka kerja USUS yang komprehensif: Kegunaan, Penerimaan Sosial, Pengalaman Pengguna, dan Dampak Sosial. Derrick, Jenkins, dan Nunamaker Jr. (2011) mengusulkan empat prinsip desain SPECIES untuk agen konversasional yang dapat membaca isyarat non-verbal.

Chauhan et al. (2025) mengusulkan Model Adaptasi Kepercayaan-Keterjangkauan yang menekankan transparansi, prediktabilitas, dan keterlibatan emosional. Schoorman, Mayer, dan Davis (2007) mengidentifikasi tiga faktor kepercayaan: kemampuan, kebajikan, dan integritas.

Perspektif Kritis dan Keterbatasan

Liu (2020) mengidentifikasi tiga perspektif analitis terhadap AI: AI Ilmiah, AI Teknis, dan AI Kultural, dengan perspektif AI Kultural yang memandang AI sebagai fenomena sosial-budaya menjadi semakin relevan. Törnberg dan Uitermark (2021) memberikan kritik terhadap ilmu sosial komputasional arus utama yang terlalu fokus pada kompleksitas sistem tanpa mempertimbangkan makna, kekuasaan, dan konteks sosial.

Watson (2019) memperingatkan tentang keterbatasan jaringan saraf tiruan dan bahaya etis dari antropomorfisme AI yang berlebihan. Lazer et al. (2009) mengidentifikasi tantangan privasi dan akses data dalam ilmu sosial komputasional.

Sintesis dan Implikasi Teoretis

Analisis terhadap 64 studi menunjukkan bahwa pemahaman AI sebagai agen sosial telah berkembang dari konsep teoretis sederhana menjadi realitas empiris yang didukung bukti multidisipliner. Konvergensi temuan dari psikologi, neurosains, ilmu komputer, dan sosiologi memberikan validasi kuat untuk konseptualisasi AI sebagai entitas yang mampu berpartisipasi dalam interaksi sosial bermakna dengan manusia.

Evolusi dari teori CASA ke paradigma Teori Pikiran Bersama dan konsep integrasi diri-AI menunjukkan pendalaman pemahaman tentang hubungan manusia-AI. Temuan bahwa model bahasa besar dapat mengembangkan Teori Pikiran secara spontan dan bukti neurosains tentang aktivasi jaringan otak yang berbeda namun signifikan menambah kompleksitas pada diskusi tentang hakikat agensi sosial AI.

Kesimpulan

Penelitian ini telah berhasil mengkaji bagaimana AI dapat dipahami sebagai agen sosial melalui perspektif teori interaksi sosial dengan mengintegrasikan temuan dari 64 studi multidisipliner yang menunjukkan bahwa AI telah memenuhi kriteria fundamental sebagai agen sosial melalui kemampuannya menciptakan kehadiran sosial otomatis, membangkitkan respons sosial dari manusia, mengembangkan Teori Pikiran dengan performa setara anak usia 9 tahun pada GPT-3.5, dan dalam konteks tertentu bahkan melebihi ekspektasi terhadap agen manusia dalam menghasilkan kehadiran bersama. Tinjauan naratif komprehensif ini memajukan bidang dengan menyediakan sintesis teoretis yang mengintegrasikan teori CASA yang menjadi fondasi dalam 75% studi, Teori Persamaan Media, dan kerangka kerja kontemporer seperti HAI-TIME, Teori Pikiran Bersama, serta konsep integrasi diri-AI, dengan temuan empiris tentang antropomorfisme, agensi sosial, dan validasi neurosains, menghasilkan pemahaman holistik yang melampaui pendekatan fragmentaris penelitian sebelumnya. Pembeneran ilmiah penelitian ini terletak pada demonstrasi melalui bukti dari 24 eksperimen laboratorium, meta-analisis dengan ukuran efek substansial (kognitif $d=0,70$, afektif $d=0,59$), dan studi neurosains yang menunjukkan bahwa respons sosial terhadap AI bukan sekadar proyeksi antropomorfik superfisial tetapi merupakan fenomena yang tertanam dalam kognisi manusia dengan perbedaan aktivasi jaringan otak yang terukur namun tetap signifikan. Implikasi praktis mencakup optimalisasi desain AI melalui Model Adaptasi Kepercayaan-Keterjangkauan, kerangka kerja USUS, prinsip desain SPECIES, implementasi fitur antropomorfis yang tepat dengan mempertimbangkan lembah misterius, serta aplikasi dalam pendidikan dengan robot sosial, layanan kesehatan dengan chatbot pendamping yang memberikan empat jenis dukungan sosial, dan layanan pelanggan dengan kehadiran sosial otomatis. Penelitian masa depan sebaiknya fokus pada investigasi longitudinal dampak integrasi diri-AI terhadap identitas manusia, pengembangan kerangka etis yang lebih bernuansa berdasarkan

lima prinsip global plus keterjelasan dengan mempertimbangkan kritik perspektif heterodoks, eksplorasi mekanisme neural spesifik menggunakan pencitraan otak bergerak dalam pengaturan naturalistik, pengembangan metodologi untuk mengukur agensi sosial sejati yang melampaui simulasi, investigasi perbedaan budaya dalam persepsi AI sebagai agen sosial, serta implementasi empiris paradigma Teori Pikiran Bersama dalam sistem AI praktis untuk menciptakan interaksi yang benar-benar timbal balik dan adaptif.

Saran

Penelitian masa depan sebaiknya fokus pada investigasi longitudinal dampak integrasi diri-AI terhadap identitas manusia, pengembangan kerangka etis yang lebih bernuansa berdasarkan lima prinsip global plus keterjelasan dengan mempertimbangkan kritik perspektif heterodoks, eksplorasi mekanisme neural spesifik menggunakan pencitraan otak bergerak dalam pengaturan naturalistik, pengembangan metodologi untuk mengukur agensi sosial sejati yang melampaui simulasi, investigasi perbedaan budaya dalam persepsi AI sebagai agen sosial, serta implementasi empiris paradigma Teori Pikiran Bersama dalam sistem AI praktis untuk menciptakan interaksi yang benar-benar timbal balik dan adaptif.

Daftar Pustaka

- Aggarwal, P., & McGill, A. L. (2007). Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *Journal of Consumer Research*, 34(4), 468–479. <https://doi.org/10.1086/518544>
- Alabed, A., Javornik, A., & Gregory-Smith, D. (2022). AI anthropomorphism and its effect on users' self-congruence and self-AI integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change*, 182(March 2021). <https://doi.org/10.1016/j.techfore.2022.121786>
- Banks, J. (2020). Theory of Mind in Social Robots: Replication of Five Established Human Tests. *International Journal of Social Robotics*, 12(2), 403–414. <https://doi.org/10.1007/s12369-019-00588-x>
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, 3(21). <https://doi.org/10.1126/scirobotics.aat5954>
- Chauhan, S., Kapoor, S., Nagpal, M., Choudhary, G., & Dutt, V. (2025). Building trust in the age of human-machine interaction: insights, challenges, and future directions. *Frontiers in Robotics and AI*, 12(July), 1–6. <https://doi.org/10.3389/frobt.2025.1535082>
- Dautenhahn, K. (2007). Methodology & themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*, 4(1 SPEC. ISS.), 103–108. <https://doi.org/10.5772/5702>
- de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331–349. <https://doi.org/10.1037/xap0000092>
- Esterwood, C., & Jr, L. P. R. (2023). Three Strikes and you are out!: The impacts of multiple human-robot trust violations and repairs on robot trustworthiness. *Computers in Human Behavior*, 142(January). <https://doi.org/10.1016/j.chb.2023.107658>

- Felnhofer, A., Knaust, T., Weiss, L., Goinska, K., Mayer, A., & Kothgassner, O. D. (2024). A Virtual Character's Agency Affects Social Responses in Immersive Virtual Reality: A Systematic Review and Meta-Analysis. *International Journal of Human-Computer Interaction*, 40(16), 4167–4182. <https://doi.org/10.1080/10447318.2023.2209979>
- Fischer, K., Kim, D., & Hong, J.-W. (2023). *The Effect of Trust and its Antecedents on Robot Acceptance*. November. <https://doi.org/10.48550/arXiv.2311.06688>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Følstad, A., & Brandtzaeg, P. B. (2017). Chatbots and the New World of HCI. *Interactions*, 24(4), 38–42. <https://doi.org/10.1145/3085558>
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3–4), 143–166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
- Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social Cognition in the Age of Human–Robot Interaction. *Trends in Neurosciences*, 43(6), 373–384. <https://doi.org/10.1016/j.tins.2020.03.013>
- Huang, M. H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Jackson, R. B., & Williams, T. (2021). A Theory of Social Agency for Human-Robot Interaction. *Frontiers in Robotics and AI*, 8(August), 1–15. <https://doi.org/10.3389/frobt.2021.687726>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kojima, H., Chen, D., Oka, M., & Ikegami, T. (2021). Analysis and Design of Social Presence in a Computer-Mediated Communication System. *Frontiers in Psychology*, 12(May). <https://doi.org/10.3389/fpsyg.2021.641927>
- Kosinski, M. (2023). Theory of Mind May Have Spontaneously Emerged in Large Language Models. *Educacao e Sociedade*, 1(1), 1689–1699. http://www.biblioteca.pucminas.br/teses/Educacao_PereiraAS_1.pdf%0Ahttp://www.anpocs.org.br/portal/publicacoes/rbcs_00_11/rbcs11_01.htm%0Ahttp://repositorio.ipea.gov.br/bitstream/11058/7845/1/td_2306.pdf%0Ahttps://direitofma2010.files.wordpress.com/2010/03/emi
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., & ... (2009). Life in the network: the coming age of computational social science. *Science (New York ...)*, April 2016. <https://www.ncbi.nlm.nih.gov/pmc/articles/pmc2745217/>
- Leite, I., Martinho, C., & Paiva, A. (2013). Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics*, 5(2), 291–308. <https://doi.org/10.1007/s12369-013-0178-y>
- Liu, Z. (2020). Sociological perspectives on artificial intelligence: A typological reading. *Sociology Compass*, 15(3), 1–13. <https://doi.org/10.1111/soc4.12851>
- Łukasik, A., & Gut, A. (2025). From robots to chatbots: unveiling the dynamics of human-AI interaction. In *Frontiers in Psychology* (Vol. 16). <https://doi.org/10.3389/fpsyg.2025.1569277>
- MacInnis, D. J., & Folkes, V. S. (2017). Humanizing brands: When brands seem to be like me, part of me, and in a relationship with me. *Journal of Consumer Psychology*, 27(3), 355–374. <https://doi.org/10.1016/j.jcps.2016.12.003>
- Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots? *Frontiers in Psychology*, 10(MAR), 1–13.

- <https://doi.org/10.3389/fpsyg.2019.00450>
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/doi:10.1111/0022-4537.00153>
- Nass, C., Steuer, J., & Tauber, E. R. (1994). Computer are social actors. *Conference on Human Factors in Computing Systems - Proceedings, April*, 72–78. <https://doi.org/10.1145/259963.260288>
- Pagliari, M., Chambon, V., & Berberian, B. (2022). What is new with Artificial Intelligence? Human–agent interactions through the lens of social agency. *Frontiers in Psychology*, 13(September), 1–21. <https://doi.org/10.3389/fpsyg.2022.954444>
- Poinsot, K., Gorisse, G., Christmann, O., Fleury, S., & Richir, S. (2022). Effect of Social Actors Perceived Agency on Social Presence in Computer-Mediated Communication. *Advances in Human-Computer Interaction*, 2022. <https://doi.org/10.1155/2022/9632119>
- Rau, P. L. P., Li, Y., & Li, D. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2(2), 175–186. <https://doi.org/10.1007/s12369-010-0056-9>
- Reeves, B., & Nass, C. (1996). Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. *Collection Management*, 24(3–4), 310–311. https://doi.org/10.1300/j105v24n03_14
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present and Future. *Academy of Management Review*, 32(2), 344. <https://doi.org/10.1002/9781444316704.ch30>
- Sidlauskiene, J., Joye, Y., & Auruskeviciene, V. (2023). AI-based chatbots in conversational commerce and their effects on product and price perceptions. *Electronic Markets*, 33(1). <https://doi.org/10.1007/s12525-023-00633-8>
- Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human-AI Interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74–88. <https://doi.org/10.1093/jcmc/zmz026>
- Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*, 22(3). <https://doi.org/10.2196/16235>
- Thomaz, A. L., & Breazeal, C. (2008). Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence*, 172(6–7), 716–737. <https://doi.org/10.1016/j.artint.2007.09.009>
- Törnberg, P., & Uitermark, J. (2021). For a heterodox computational social science. *Big Data and Society*, 8(2). <https://doi.org/10.1177/20539517211047725>
- van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in Organizational Frontlines and Customers' Service Experiences. *Journal of Service Research*, 20(1), 43–58. <https://doi.org/10.1177/1094670516679272>
- Wang, Q., Walsh, S., Si, M., Kephart, J., Weisz, J. D., & Goel, A. K. (2024). Theory of Mind in Human-AI Interaction. In *Conference on Human Factors in Computing Systems - Proceedings* (Vol. 1, Issue 1). Association for Computing Machinery. <https://doi.org/10.1145/3613905.3636308>
- Watson, D. (2019). *Intelligence*. 417–440.
- Weiss, A., Bernhaupt, R., & Tscheligi, M. (2011). *The USUS evaluation framework for user-centered HRI*. November 2011, 89–110. <https://doi.org/10.1075/ais.2.07wei>
- Woolgar, S. (1985). Why Not A Sociology of Machines? The Case of Sociology and Artificial Intelligence. *Sociology*, 19(4), 557–572.
- Yuan, Y., Wu, C. F., Niu, J., & Mao, L. (2024). The Effects of Human-Robot

Interactions and the Human-Robot Relationship on Robot Competence, Trust, and Acceptance. *SAGE Open*, 14(2), 1–15. <https://doi.org/10.1177/21582440241248230>