



Pengembangan Konsep Ekosistem AI Mandiri untuk Multi-Model berbasis Msty

Christopel H. Simanjuntak¹, Arief P. Kumaat², Adelaida Joroh³,
Stieven N. Rumokoy⁴, Deby A. Mpila⁵, Adrian Warokka⁶

¹Program Studi Teknik Informatika, Jurusan Teknik Elektro, Politeknik Negeri Manado,
Manado

²Program Studi Manajemen Pemasaran, Jurusan Administrasi Bisnis, Politeknik Negeri Manado,
Manado

³Program Studi Manajemen Bisnis, Jurusan Administrasi Bisnis, Politeknik Negeri Manado,
Manado

⁴Program Studi Teknik Listrik, Jurusan Elektro, Politeknik Negeri Manado,
Manado

⁵Program Studi Farmasi, Fakultas Matematika dan Ilmu Alam, Universitas Sam Ratulangi,
Manado

⁶Program Studi Teknik Mesin Produksi dan Perawatan, Jurusan Teknik Mesin, Politeknik Negeri
Manado, Manado

E-mail: christopel.simanjuntak@polimdo.ac.id

Abstrak

Infrastruktur cloud pihak ketiga menuntut pengembangan Artificial Intelligence (AI), tetapi sering kali hal ini disertai kekhawatiran privasi, pelanggaran kedaulatan data, dan biaya operasional yang mahal. Namun demikian, belakangan ini, paradigma pengembangan AI lokal (on-premise) telah mendapatkan daya tarik sebagai alternatif yang lebih baik. Makalah ini membahas gagasan dan fondasi pembangunan ekosistem AI otonom untuk lingkungan multi-model di Msty. Metode penelitian adalah studi konseptual dan pendekatan eksperimental untuk menguji ruang kerja desktop terintegrasi seperti itu dari model bahasa besar untuk pembuatan teks dan model lokal untuk menghasilkan gambar. Hasil penelitian menunjukkan bahwa Msty dapat mengoordinasikan interaksi multi-model secara bersamaan melalui metode manajemen memori internal yang efisien tanpa ketergantungan komputasi eksternal. Menerapkan ekosistem ini pada akhirnya mencapai kedaulatan data penuh dan fleksibilitas dalam alur kerja bagi pengembang. Penelitian ini menyimpulkan bahwa Msty efektif mempercepat demokratisasi AI mandiri, sekaligus menjadi aplikasi tepat untuk pengembangan kecerdasan buatan yang aman, privat, dan hemat biaya pada perangkat keras tingkat konsumen.

Kata kunci— Local-Center; Ekosistem Mandiri; Multi-Model; Msty.

1. PENDAHULUAN

Dalam beberapa tahun terakhir, teknologi Kecerdasan Buatan (AI) telah membuat lompatan besar untuk mendefinisikan ulang paradigma komputasi global melalui Model Bahasa Besar (LLM) dan model generator visual [1]. Tidak hanya komputasi cerdas ini lagi-lagi bukan sekadar teknologi dasar yang statis, tetapi telah berkembang menjadi kekuatan utama bagi efisiensi alur kerja di bidang kreatif, akademik, dan perusahaan besar. Namun, bahkan hingga

saat ini, banyak adopsi AI masih bergantung pada infrastruktur cloud yang dikendalikan oleh penyedia layanan pihak ketiga yang bersifat monolitik. [2]. Demi meminimalisir risiko privasi data, menyingkirkan privasi data yang dirisiki oleh pengaturan pihak ketiga, dan biaya operasional berlangganan jangka panjang, terdapat tiga tantangan krusial bagi pengembang, peneliti, dan organisasi yang muncul dari ketergantungan kronis pada server terpusat ini [3]. Ketika data yang sensitif dikirim ke server luar, resiko data yang bocor menjadi ancaman yang nyata dan sangat sulit dihindari.

Sebagai reaksi terhadap tantangan tersebut, paradigma Local AI atau On-Premise AI mulai muncul sebagai alternatif yang menjanjikan di kalangan komunitas pengembang open-source[4]. Pemanfaatan model AI secara lokal ke perangkat keras secara lokal memastikan data tidak pernah keluar dari ekosistem fisik komputer pengguna, dan menawarkan jaminan keamanan dan kerahasiaan tinggi [5]. Menawarkan privasi mutlak, keuntungan utama dari pemanfaatan AI lokal ini adalah hambatan masuk yang tinggi (barrier to entry) karena kompleksitas konfigurasi teknis. Meskipun proses instalasi pustaka pemrograman yang rumit, masalah manajemen dependensi pustaka system, dan masalah alokasi memori sering menjadi dinding penghalang yang besar [6]. Membangun alur kerja terpadu dan harmonis yang mendorong pengembangan lebih dari satu model AI (multi-model), seperti model teks berbasis bahasa dan model generator gambar berbasis visual, menjadi semakin kompleks [7].

Dalam konteks fragmentasi perangkat lunak inilah platform Msty [8] hadir sebagai solusi konseptual dan praktis untuk menjembatani hambatan teknis tersebut. Msty menawarkan sistem manajemen desktop inovatif yang dirancang khusus untuk mengorkestrasi ekosistem multi-model secara lokal dengan efisiensi memori yang tinggi [9]. Berbeda dengan penginstal berbasis kontainer tradisional yang cenderung berat, memakan ruang penyimpanan masif, dan menuntut spesifikasi peladen (server). Msty [10] mengedepankan integrasi alur kerja yang ringkas dan adaptif pada perangkat keras konsumen. Platform ini mengonseptualisasikan ulang bagaimana interaksi antara pengguna, model bahasa besar, dan generator gambar dapat berjalan di dalam satu ruang kerja tunggal tanpa memicu kegagalan memori sistem (*system crash*).

Dengan latar belakang tersebut, penelitian ini bertujuan untuk mengevaluasi pendekatan konseptual dalam membangun ekosistem AI lokal yang mandiri [11][12]serta menguji kemampuan orkestrasi platform Msty dalam menyatukan model teks dan gambar tanpa ketergantungan pada komputasi luar[13]. Bagaimana Msty mengelola koeksistensi multi-model pada komputer berspesifikasi menengah menjadi fokus penelitian ini. Secara teoretis, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam memperkaya literatur ilmiah desentralisasi AI, komputasi tepi (*edge computing*), dan kedaulatan data. Secara praktis, artikel ini diproyeksikan menjadi cetak biru (*blueprint*) aplikatif bagi para peneliti, akademisi, dan praktisi teknologi untuk membangun lingkungan kerja kecerdasan buatan yang aman, privat, fleksibel, dan hemat biaya secara mandiri.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah pendekatan kualitatif-eksploratif dengan metode rekayasa konseptual (conceptual engineering) dan pemodelan arsitektur sistem. Fokus utama dari metode ini adalah desain, struktur, dan evaluasi framework konseptual untuk sebuah ekosistem AI mandiri yang dapat menampung beberapa model secara sekaligus (multi-model) dan dapat berjalan lokal. Tidak untuk menuji hipotesis statistik, penelitian penyelidikan di sini ditujukan untuk konseptualisasikan sebuah paradigma baru dalam implementasi AI on-premise yang aman, efisien, terdesentralisasi dengan memanfaatkan platform Msty sebagai objek kajian utama. Tahapan penelitian dimulai dengan Analisis Kebutuhan Konseptual yang

dilakukan dengan studi literatur yang mendalam mengenai kedaulatan data (*data sovereignty*), desentralisasi AI, dan multi-modality. Di sini, hambatan-hambatan teknis utama (seperti fragmentasi dependensi *software* dan VRAM terbatas) yang selama ini menghalangi adopsi AI lokal oleh pengguna akhir diidentifikasi. Dari analisis yang telah dilakukan dirumuskan prinsip-prinsip dasar yang harus dipenuhi oleh sebuah ekosistem AI mandiri agar dapat berjalan optimal pada perangkat keras tingkat konsumen.

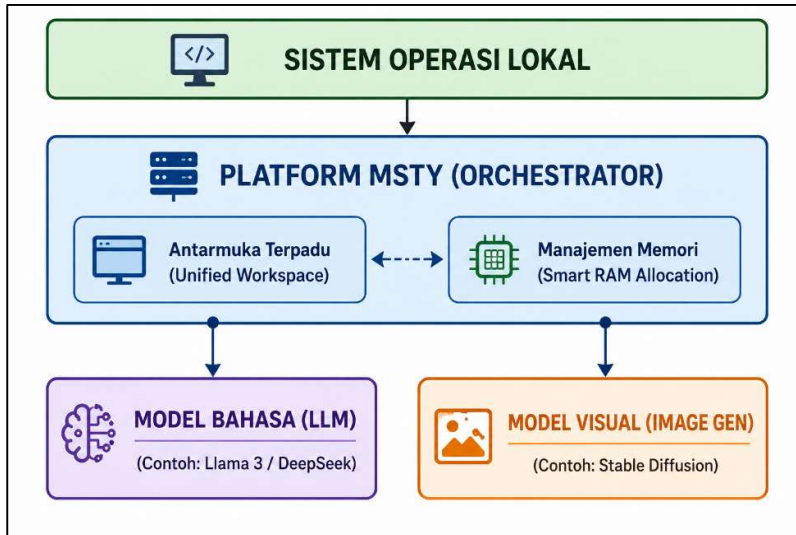
Tahap kedua adalah Perancangan Arsitektur Sistem dan Alur Kerja (*Workflow Modeling*). Peneliti menyusun model abstrak[14] yang menggambarkan bagaimana platform Msty memfasilitasi koeksistensi antara model bahasa besar (*Large Language Models*) [15] dan model generator visual (seperti *Stable Diffusion*). Fokus dalam perancangan ini adalah pada 3 konsep utama, *Unsified Interface Concept*, yaitu menyatukan ruang kerja teks dan gambar dalam satu antarmuka; *Dynamic Resource Allocation Concept*, yaitu konsep ketika model menghentikan dan mengaktifkan secara bergantian di sistem untuk menghemat RAM; *Isolated Data Environment*, yaitu seluruh data pemrosesan luring dan privat.

Tahap ketiga adalah Uji Konseptual (*Conceptual Validation*), yang dilakukan dengan simulasi skenario alur kerja pada perangkat keras dengan spesifikasi rendah. The purpose of this simulation is to validate whether it is possible to implement the model design in real-life without causing the system to fail (system crash). Evaluasi dilakukan dengan mengevaluasi kelayakan alur kerja operasional, kemudahan integrasi antar-model, dan efektivitas manajemen memori yang ditawarkan oleh Msty. Ketika itu selanjutnya dilakukan Penarikan Kesimpulan. Data kualitatif dan hasil pengamatan arsitektur hasil simulasi digabungkan untuk merumuskan sebuah cetak biru (blueprint) formal. Selanjutnya, cetak biru ini disampaikan sebagai sumber teoritis baru bagi peneliti dan praktisi pengembangan AI lokal di masa depan.

3. HASIL DAN PEMBAHASAN

3.1 Konseptual Ekosistem Local-Centre AI secara Mandiri

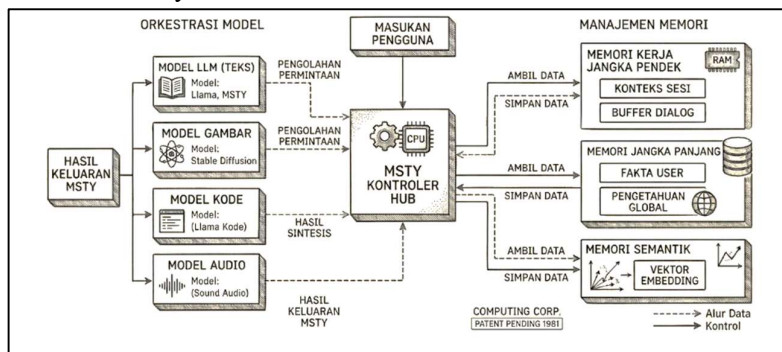
Hasil desain menunjukkan, pergeseran paradigma dari arsitektur cloud terpusat ke arsitektur edge-computing yang terdesentralisasi diperlukan dalam pengembangan ekosistem AI secara independen (*self-contained*) pada perangkat keras konsumen. Sasaran yang ditetapkan dalam penelitian ini adalah konsep ekosistem mandiri yang dibangun di atas tiga pilar utama, yaitu kedaulatan data mutlak (*absolute data sovereignty*), interoperabilitas multi-model luring, dan efisiensi sumber daya berbasis kebutuhan (*on-demand resource efficiency*). Dalam skema konseptual ini, semua data yang digunakan untuk memproses, dari prompt input hingga inferensi model dan generasi output visual dan tekstual, terdapat dalam memori lokal dari user. Dengan konsep ini secara teoretis, kelemahan yang paling umum di layanan AI berbasis cloud API, yaitu *man in the middle* (MITM) dapat dieliminasi.



Gambar 1. Skema Konseptual Sistem

3.2 Orkestrasi Multi-Model dan Manajemen Memori pada Msty

Implementasi konsep multi-model co-existence (koeksistensi multi-model) menghadapi tantangan besar pada perangkat keras dengan batasan RAM 16GB. Menjalankan Model Bahasa Besar (LLM) seperti Llama 3 (8B) secara bersamaan dengan model generator bahasa seperti Qwen secara teoretis akan melampaui kapasitas memori fisik sistem. Namun, Msty melompati melebihi kebuntuan konseptual ini dengan melakukan pendekatan *Dynamic Memory Swapping* (Pertukaran Memori Dinamis). Msty bertindak sebagai pengatur lalu lintas (orchestrator) cerdas yang mengalokasikan RAM secara bergantian berdasarkan fokus tugas pengguna. Ketika pengguna berinteraksi dengan Chatbot Teks, Msty akan prioritaskan seluruh memori yang tersedia untuk memuat Bobot (weights) dari LLM. Begitu pengguna beralih ke instruksi pembuatan gambar, Msty secara otomatis mengistirahatkan proses LLM ke dalam memori virtual (swap) dan mengaktifkan model visual secara instan. Manajemen memori hibrida ini cukup efektif untuk mencegah terjadinya kegagalan sistem (*system crash atau Out of Memory Error*) pada prosesor i5 dan RAM 16GB, sekaligus menjaga komputer tetap responsif untuk tugas komputasi dasar lainnya.

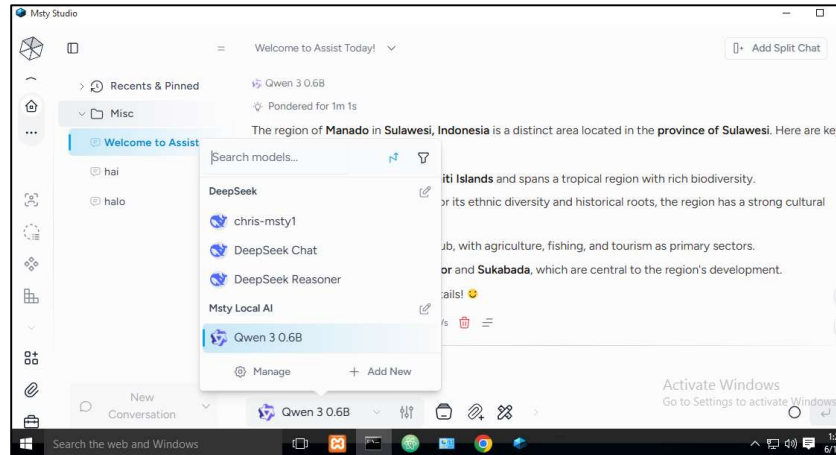


Gambar 2. Sistem Kerja Msty.

3.3 Unifikasi Antarmuka dan Fleksibilitas Alur Kerja

Secara arsitektur perangkat lunak, Msty berhasil menghancurkan dinding pemisah antarmuka yang selama ini menjadi kelemahan utama platform open-source tradisional, seperti Pinokio. Pinokio memaksa pengguna untuk mengelola beberapa jendela *command prompt* dan

lingkungan virtual terisolasi (*Python venv/Docker*) untuk setiap aplikasi AI individual. Sebaliknya, Msty menggunakan konsep Arsitektur Ruang Kerja Terpadu (*Unified Workspace Architecture*). Dalam satu aplikasi desktop Msty, pengguna dapat mengunduh, mengonfigurasi, dan menjalankan model teks serta gambar dalam satu garis waktu obrolan (*chat timeline*). Dengan integrasi ini, tidak hanya menghapus hambatan teknis bagi pengguna awam, tetapi juga membangun alur kerja (*workflow*) yang sangat efisien bagi pengembang. Pengguna dapat memberikan perintah kepada LLM lokal untuk membuat deskripsi prompt yang detail.



Gambar 3. Tampilan Unifikasi Msty dengan beberapa Model di Lokal Komputer.

3.4 Keunggulan Paradigma AI Lokal

Kajian ini menunjukkan keunggulan yang jauh dari Msty dengan kompetitornya. Dibandingkan dengan Jan AI yang masih sangat memusatkan dan mengutamakan fungsionalitas teks, Msty satu langkah lebih maju dalam proses matang konsep multi-modality visual. Setelah dibandingkan dengan Pinokio, Msty lebih unggul dalam hal kesederhanaan operasional. Msty menghapus kebutuhan ruang puluhan gigabyte untuk dependensi sistem yang tumpang tindih, menjadikannya cetak biru terbaik untuk demokratisasi AI mandiri. Ketergantungan pada server besar global ini terbukti dapat dipangkas secara signifikan tanpa terhenti pada fungsionalitas teknologi kecerdasan buatan modern secara utama..

4. KESIMPULAN DAN FUTURE WORK

Penelitian ini berhasil merumuskan kerangka konseptual pengembangan ekosistem AI mandiri untuk lingkungan multi-model lokal dengan menggunakan platform Msty. Mengecilkan ketergantungan kronis terhadap infrastruktur cloud pihak dengan pendekatan komputasi tepi (*edge computing*) terdesentralisasi, berdasarkan hasil analisis arsitektur. Platform Msty tepat mengatasi hambatan teknis yang ada pada platform pendahulunya seperti Pinokio melalui penerapan *Unified Workspace Architecture*. Walaupun perangkat keras konsumen terbatas, Msty dapat menyelesaikannya secara cerdas dengan *Dynamic Memory Swapping*. Sistem ini dapat mengkoordinasikan model bahasa besar dan model generator visual secara alternatif tanpa menyebabkan kegagalan sistem (*system crash*). Dengan demikian, ekosistem mandiri ini memberikan pengembang fleksibilitas alur kerja yang tinggi dan merekauliatisasi kedaulatan data dan privasi pengguna akhir di dalam biaya yang efektif.

Pembahasan kuantitatif dalam penelitian ini masih ada batasannya. Untuk itu, penelitian kedepan adalah analisis empiris dan eksperimen pengukuran metrik performa yang dilakukan secara eksak pada beberapa variasi kartu grafis (GPU), yaitu: Latensi generasi per token,

Kecepatan pemrosesan gambar per detik (it/s), Fluktuasi konsumsi daya listrik. Keberadaan penelitian masa depan juga diharapkan untuk mengeksplorasi integrasi konsep ini dalam jaringan lokal skala kecil (LAN) di sektor pendidikan atau korporasi untuk menguji efektivitas kolaborasi multi-user di ekosistem AI lokal yang terdesentralisasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Politeknik Negeri Manado atas dukungannya dalam memberikan ruang agar tulisan ini bisa diseminarkan dan dipublikasi.

DAFTAR PUSTAKA

- [1] M. Alirezaie, W. Hoffman, P. Zabihi, H. Rahnama, and A. Pentland, "Decentralized Data and Artificial Intelligence Orchestration for Transparent and Efficient Small and Medium-Sized Enterprises Trade Financing," 2024.
- [2] M. L. M. Multi-llm, "Toward Edge General Intelligence with Multiple-Large Language Model (Multi-LLM): Architecture, Trust, and Orchestration," *IEEE Trans. Cogn. Commun. Netw.*, pp. 1–23, 2025.
- [3] A. F. T. Winfield, "Experiments in Artificial Theory of Mind: From Safety to Story-Telling," *J. Front. Robot. AI*, vol. 5, no. June, 2018.
- [4] CloudStack-LLC, "Local Models - Msty Studio Docs," *Msty*, 2025. [Online]. Available: <https://docs.msty.studio/managing-models/local-models>. [Accessed: 30-May-2026].
- [5] X. Shen, D. J. Wagg, M. Tipuric, and M. S. Bonney, "Digital twins as self-models for intelligent structures," *Sci. Rep.*, pp. 1–11, 2025.
- [6] S. Somanathan, "Orchestrating Agentic AI in Program Management: Governance Frameworks for Autonomous Project Workflows," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 4, pp. 193–200, 2025.
- [7] A. Zafeiropoulos, N. Filinis, E. Fotopoulou, and S. Papavassiliou, "AI-Assisted Synergetic Orchestration Mechanisms for Autoscaling in Computing Continuum Systems," *IEEE Commun. Mag.*, vol. 63, no. 1, pp. 116–122, 2025.
- [8] CloudStack-LLC, "Unleashing Local AI - Transform Your Coding Experience with Msty | msty.ai," *Msty*, 2026. [Online]. Available: <https://msty.ai/blog/local-ai-coding-assistants/>. [Accessed: 30-May-2026].
- [9] D. C. Youvan, "AI Hosting AI-Me and Me: Recursive Instantiation, Logogenic Resonance, and the Emergence of the Personal Robot," pp. 1–30, 2025.
- [10] CloudStack-LLC, "Msty Studio | AI Workspace for Chat, Media, and Automation | msty.ai," *Msty*, 2026. [Online]. Available: <https://msty.ai/studio/>. [Accessed: 30-May-2026].
- [11] O. Issn, P. Issn, and I. J. A. Res, "AI NEXUS 910+: A MULTI-AGENT ORCHESTRATED UNIFIED AI TOOL PLATFORM WITH WORKFLOW AUTOMATION," *Int. J. Adv. Res.*, vol. 14, no. 3, pp. 842–848, 2026.
- [12] Y. Zeng, Y. Zhao, J. Bai, and B. Xu, "Toward Robot Self-Consciousness (II): Brain-Inspired Robot Bodily Self Model for Self-Recognition," *Cognit. Comput.*, vol. 10, no. 2, pp. 307–320, 2018.
- [13] V. T. Ponnaganti, "Scalable Multi-Model Orchestration in AI Microservices with Kubernetes and Serverless for Event-Driven MLOps Pipelines," in *2025 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2025, pp. 1471–1476.
- [14] D. Khati, Y. Liu, D. N. Palacio, Y. Zhang, and D. Poshyvanyk, "Mapping the Trust Terrain: LLMs in Software Engineering - Insights and Perspectives," *ACM Trans. Softw. Eng. Methodol.*, vol. 1, no. 1, pp. 1–38, 2025.
- [15] D. J. Gunkel, "Hands-On Large Language Models Language. Understanding and Generation," in *Ethos, Technology, and AI in Contemporary Society*, First Edit., N. Butterfiled, Ed. United States: O'Reilly, 2024, pp. 231–248.
