ENGELBERT: COMPLETING ENGELBART'S REVOLUTION—A SOVEREIGN OS FOR AUGMENTED THOUGHT *

David Angaya
Independent Researcher
Nairobi, Kenya
davidomungala3@gmail.com

Davin Dewanto
Ludwig Maximillian Universitat
Munich, Germany
Davindewanto@gmail.com

ABSTRACT

This paper presents a vision for the next generation of computing, centered on an operating system architecture that inverts the reactive GUI paradigm. We propose **Engelbert**, a sovereign OS designed to function as a Socratic partner in thought. Its flagship environment, **Wise**, is the first integrated thinking environment combining a local "Second Brain," a proactive "Insight Engine," and adaptive "Cognitive Lenses"—all with local-first execution. **We introduce a novel distribution model for the initial phase using 'llama-cpp-python' to create a robust, hardware-aware user experience, and a zero-knowledge encryption architecture for optional cloud synchronization.** This vision paper outlines Engelbert's architecture, introduces Wise as the first step on a roadmap from portable to ambient intelligence, and proposes a research agenda for the future of **sovereign, ubiquitous, and ultimately post-screen cognitive computing**.

1 Introduction

For half a century, the Graphical User Interface (GUI) has defined human-computer interaction, casting the user in the role of a "low-level system operator" (Engelbart, 1962). This paradigm, while revolutionary, creates a bottleneck for complex thought. Our work is a deliberate counter-movement to the "Attention Economy" (Goldhaber, 1997); instead of building another system to capture attention, we propose an "attention shield" designed to amplify and direct the user's own cognitive resources.

We believe the rise of capable local models allows us to finally complete Douglas Engelbart's unfinished revolution. EngelBERT is not merely a successor to today's OSes, but a fundamental rethinking of human-computer collaboration. This paper outlines the architecture for this new paradigm. We are not building an application; we are laying the first stone of a **cathedral** for the mind. The Wise Sovereign MVP, currently in development, is that first stone—a tangible proof of the grander vision.

Our goal is not to replace task automation, but to build upon it. While powerful "AI Butlers" like Microsoft Copilot excel at reducing cognitive burden, they stop short of augmenting the user's own reasoning process. EngelBERT uses task automation as raw material for a deeper, Socratic partnership. Through its flagship application, **Wise**, users interact with a system that listens, anticipates, and debates, bridging the gap between automation and true cognitive augmentation. Our architecture resolves a key tension in modern AI systems by pioneering a hybrid-sovereign model. The system is local-first by default, but supports optional, end-to-end, zero-knowledge encrypted synchronization of the user's "Second Brain" via a service we call "Wise Cloud." This provides the absolute privacy required for a true thought partner with the convenience of a modern, multi-device workflow.

2 Theoretical Foundation

Our vision is grounded in a synthesis of cognitive science, historical computing visions, and emerging trends in AI.

* *Citation*: **Angaya, D., & Dewanto, D. (2025). EngelBERT: A Vision for a Socratic Partner in Thought.**

2.1 Intellectual Legacy and Positioning

EngelBERT unifies several historical visions into a single, cohesive architecture. It seeks to realize the "intelligence amplification" of Engelbart, (Engelbart, 1962), the "man-computer symbiosis" of Licklider, (Licklider, 2021)** on a local and sovereign basis, and the "associative memory" of Bush, (Bush et al., 1945) Memex through its proactive Insight Engine.Its flagship application, Wise, in turn, realizes Alan Kay's (Kay, 1996) vision for the Dynabook as a dynamic, personal learning companion.

2.2 Architectural Positioning

Our approach represents a distinct path toward an "AI-OS." While some research explores replacing low-level components like schedulers with LLMs (Miehling et al., 2025), EngelBERT is architected as an **application-level "meta" OS**. We do not seek to replace the stable, underlying operating system (like Windows or macOS). Instead, we provide a new, intelligent, and intent-driven interaction layer that sits on top, leveraging the stability of existing infrastructure. To our knowledge, no existing system is architected around the principles of proactive, local-first, Socratic partnership for the express purpose of cognitive augmentation.

2.3 Cognitive Science and Human-AI Augmentation

EngelBERT is inspired by the parallel processing and distinct memory systems of the human brain. The dual process theory of cognition (Kahneman, 2011), which distinguishes between fast, intuitive "System 1" thinking and slow, analytical "System 2" thinking, informs our architectural design. We aim to create a system that can fluidly support both cognitive modes, augmenting human intellect rather than simply automating tasks. This aligns with a growing body of research exploring how humans and AI can collaborate to improve outcomes (Miehling et al., 2025). Furthermore, our core mission is grounded in the formal study of Human Cognitive Augmentation. Research in this field proposes that a human-AI partnership should be measured not just by efficiency, but by its ability to increase "Cognitive Accuracy" (producing a desired result) and "Cognitive Precision" (avoiding undesired results). Our Cognitive Lenses are a direct architectural attempt to build a system that can be optimized for these metrics (Fulbright, 2019).

2.4 Proactive Computing and Ambient Intelligence

The core philosophy of our proactive "Insight Engine" builds on decades of research in Proactive Computing and Ambient Intelligence (Aarts & De Ruyter, 2009). Researchers have long envisioned "intelligent environments" that are context-aware and anticipatory (Dey, 2001). We argue that the advent of powerful LLMs provides the missing "cognitive engine" critical for realizing this long-standing vision at scale.

2.5 Positioning Within the Agentic AI Landscape

Frameworks like LangChain (Chase, 2022), Microsoft's Semantic Kernel, and autonomous systems like AutoGPT have been pivotal in popularizing agentic workflows. However, they remain primarily reactive toolkits designed to help developers orchestrate responses to user commands. EngelBERT proposes a higher level of abstraction: a prototype kernel evolving toward an OS for agentic computing. Our architecture's key differentiator is the integration of a proactive synthesis loop (the Insight Engine) and a mechanism for user-directed reasoning (Cognitive Lenses)—a philosophical shift from reactive task automation to proactive cognitive augmentation.While adjacent tools like Google's NotebookLM have begun to apply AI to personal documents, they remain fundamentally reactive and cloud-dependent.

3 The EngelBERT OS Vision

Our architecture is composed of three concentric layers: a central Cognitive Core, an Augmented Workspace (Wise), and a future Agent Economy.

3.1 A Day in the Life with Wise

The Wise application (powered by Engelbert) is the first environment to unify a sovereign "Second Brain," agentic tools, and proactive reasoning.

Consider Sabina, a developer using our flagship application, Wise. She has been documenting her project for weeks, storing code snippets and design notes in her local "Second Brain." After pushing a new feature, she gets a bug report.

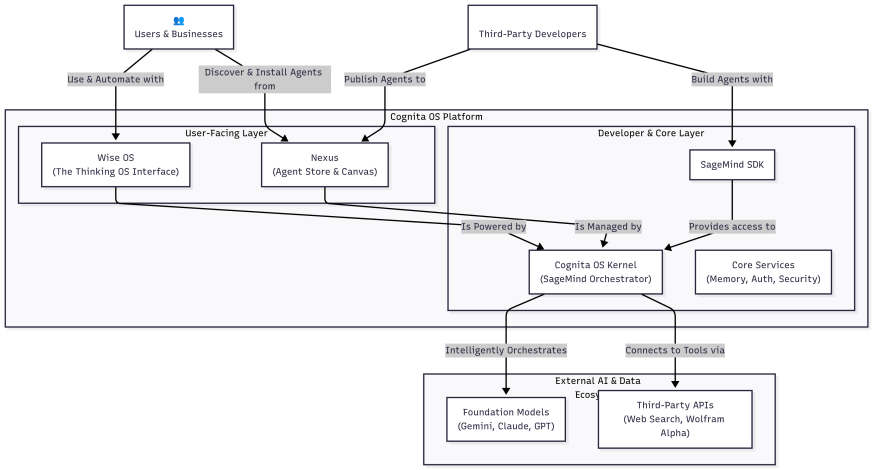

Figure 1: The EngelBERT Ecosystem, with the Kernel, Workspace, and a future Agent Economy.

She asks Wise: "Review my recent commits related to the user authentication flow and my original design notes from last month."

Later, while she is debugging, the Insight Engine, running its passive 'dream cycle' on her local vector store, surfaces a non-intrusive suggestion: "I noticed the logic in your handleAuthRedirect function conflicts with a constraint you mentioned in a design note from three weeks ago: 'Ensure all redirects are server-side validated to prevent open-redirect vulnerabilities.' This might be the source of the bug." Intrigued, Sabina activates the "Agora" Cognitive Lens, loading a specific master prompt and toolset into the DPES loop to debate the architectural trade-offs of the fix, then switches to the "Muse" lens to brainstorm a more elegant solution.

3.2 The Long-Term Vision: An Intent-Driven Cognitive Operating System

The Wise application prototypes our underlying technology. The full EngelBERT OS architecture is designed to be a true platform.

- **The Engelbert Kernel:** A master Tool Router using a DPES loop to deconstruct user intent and dispatch tasks to the best available agent.
- **Agentic Tool Use:** A planned capability where the orchestrator chains agents and tools, enabling complex, multi-step workflows from a single natural language command.
- **The Nexus Agent Economy & Knowledge Refinery:** Our ultimate goal is a future marketplace where developers can build and sell specialized agents that plug into EngelBERT. This creates an open economy for cognitive tools. Furthermore, this lays the foundation for a **"Knowledge Refinery"**—an ethical, user-controlled data marketplace. Users can, with explicit consent, contribute anonymized insights from their "Second Brain" to a global pool, training ever-smarter specialized models and receiving a share of the value created.

3.3 From Portable to Ambient Intelligence: The Path to a Post-Screen Future

The GUI is a bottleneck. Our ultimate vision is to pioneer what comes after. The Engelbert OS is architected not for a single device, but as an ambient, personal intelligence layer that follows the user. We see a two-phase evolution.

Phase 1: Portable Intelligence (The Sage Stick). Our initial implementation, the Wise application, embodies the principle of *portable intelligence*. It is a self-contained, sovereign brain that a user can carry with them on a device like a laptop or a USB drive (the "Sage Stick"). It plugs into existing "dumb" terminals (any computer) and temporarily endows them with the user's personal cognitive partner. The screen is still necessary in this phase, primarily as a trust-building interface that makes the AI's invisible thinking process visible via our "Cognitive Cockpit" UI.

Phase 2: Ambient Intelligence (The Smart Radio). This is the cathedral. In this future, Engelbert graduates from an application to a true ambient OS. It does not live on a device; it is a persistent, intelligent presence. Its primary interface is not a screen, but voice—a **"Smart Radio"** that is always with you, a Socratic partner made manifest. This

sovereign body has its own compute and connectivity. It discovers and orchestrates other devices in the environment. It doesn't just "control the fridge"; it understands the user's high-level intent ("I'm tired and have a big presentation tomorrow") and coordinates a network of ambient agents to serve that intent. This is the ultimate "attention shield": a future where the screen is killed entirely, freeing human consciousness from the tyranny of the glowing rectangle.

4 Core Architecture and Technical Innovations

4.1 The DPES Loop: An Intent-Driven Orchestration Engine

The Cognitive Core operates on a Deconstruct-Plan-Execute-Synthesize (DPES) loop. A fine-tuned classification model deconstructs unstructured user intent into a structured plan. This loop includes a "Critic Agent" for self-correction—a novel architectural pattern where a second LLM evaluates the primary response for quality, triggering an automatic escalation to a more powerful reasoning chain if the initial response is insufficient.

Engelbert's SageMind Orchestrator also handles deterministic tasks (e.g., calendar scheduling, data retrieval) via specialized agents—not as an end goal, but to free cognitive bandwidth for the Insight Engine's proactive synthesis

Algorithm 1 The DPES Loop with Self-Correction

```
procedure ORCHESTRATE(user_prompt, active_lens)
                                                                  ▷ 1. Deconstruct
    plan ← DeconstructIntent(user_prompt)
    execution_graph ← CreateExecutionPlan(plan)
                                                                  ▷ 2. Plan & 3. Execute
    results ← {}
    for step in execution_graph.steps_in_order() do
        input_data ← GetInputsForStep(step, results)
        primary_result ← step.agent.run(input_data, lens=active_lens)
                                                                  ▷ Self-correction via Critic Agent
        critic_score ← CriticAgent.evaluate(primary_result)
        if critic_score < THRESHOLD then
                                                                  ▷ Escalate to a more powerful agent or RAG pipeline
            final_result ← step.escalation_agent.run(input_data, lens=active_lens)
        else
            final_result ← primary_result
        end if
        results[step.id] ← final_result
    end for
                                                                  ▷ 4. Synthesize
    return SynthesizeResults(results, user_prompt)
end procedure
```

4.2 The Insight Engine: A Computational "Dream/Drift Cycle"

The human brain's ability to generate novel ideas often comes not from focused effort, but from the mind's tendency to "drift"—making loose, associative connections between disparate memories. Our Insight Engine is architected to be a computational "dream/drift cycle" for the user's "Second Brain." It's a background process: The engine runs constantly and passively, requiring no active input from the user. It works with the user's "memories": It operates on the user's entire local knowledge base—conversations, uploaded documents, and saved notes. It finds non-obvious connections: Using graph analysis and clustering algorithms on the data's embeddings, it searches for semantically related concepts that have not been explicitly linked by the user. It produces "eureka" moments: When a high-potential connection is found, it is surfaced as a non-intrusive suggestion in the "Insights Feed" (e.g., "I noticed a connection between your notes on 'X' and your conversation about 'Y'...").

While engineering true serendipity remains a frontier challenge, our architecture is designed to manage the signal-to-noise ratio through a "Focus Score," a local measure of user activity that determines when to surface insights. Insights are queued during high activity and presented more directly during idle time, creating an empathetic, non-disruptive interface.

4.3 Cognitive Lenses: User-Directed Reasoning

Cognitive Lenses" are configuration objects that modify the OS's reasoning style by modifying the 'Execute' step of the DPES loop. Each lens contains a specific system prompt, model parameters, and curated tools:

- **Scholar (Propositional):** For rigorous, analytical thinking.
- **Muse (Participatory):** For creative, divergent brainstorming.
- **Reflective (Perspectival):** For metacognitive self-inquiry. Inspired by cognitive behavioral therapy, it detects cognitive distortions and prompts comparisons with past context to foster meta-cognitive growth.
- **Agora (Procedural):** For collaborative, task-oriented dialogue.

5 Implementation and Evaluation

5.1 Towards Implementation: A Sovereign Appliance Model

Our MVP, the "Wise Sovereign Edition," represents the first phase of our vision: **Portable Intelligence**.

Our core technical innovation for distribution is a hardware-aware kernel that uses 'llama-cpp-python'. On first launch, this library compiles a C++ inference engine optimized for the user's specific CPU architecture. This solves the critical "works on my machine" problem of pre-compiled binaries and ensures a robust, reliable text-based experience even on low-spec hardware without modern CPU features.

This architecture enables a "Portable Mode," where the entire application and the user's "Second Brain" database can be run from a portable USB drive—the **"Sage Stick."** This allows a user to carry their entire cognitive partner with them, leaving zero trace on a host machine. The desktop app is designed to be self-contained for core text functions, but also implements **"Progressive Sovereignty"**: it can intelligently detect a local Ollama installation to unlock more powerful, multimodal features for users with capable hardware.

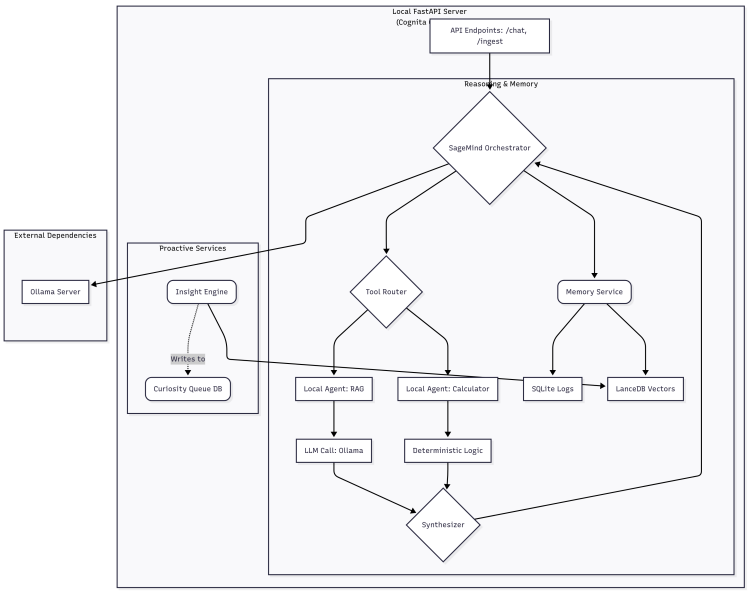

Figure 2: A conceptual mockup of the "Wise Sovereign Scholar Edition" Architecture.

5.2 A Commitment to Empirical Validation

We propose a framework for empirical validation, grounded in the formal metrics of cognitive augmentation that future work should adopt. Our immediate next step is a formal user study (N=10) designed to measure the increase in Cognitive Accuracy and Precision (Fulbright, 2019) provided by our system, compared against a standard text editor as a control.

- To measure **Cognitive Accuracy**, we will test our Muse Lens on a creative task, quantifying the increase in novel ideas generated (an "Idea Diversity Score").
- To measure **Cognitive Precision**, we will test our Scholar Lens on an analytical task, quantifying the user's ability to identify factual inconsistencies in a document (a "Solution Robustness Score").

Future implementations should assess cognitive load (via NASA-TLX) to validate the cognitive augmentation hypothesis.

6 Challenges and a Research Agenda

A credible vision must face its greatest obstacles.

- **Reliability and Safety:** LLM reliability remains a primary blocker. Our mitigation strategy combines architectural solutions (the Critic Agent) with user-centric design. To prevent over-reliance and filter bubbles, the system is designed to be Socratic, featuring cognitive fixation detection and transparent provenance for all insights.
- **The Sovereignty-Intelligence Trade-off:** We are exploring hybrid models that use local-first execution by default, with transparent, user-consented escalation to cloud APIs for complex tasks.The question remains, what is the optimal architecture for hybrid models that balance local privacy with cloud power?
- **The Cognitive Load Distribution Problem:** A key research challenge is developing a principled science for partitioning cognitive work between humans and AI to maximize augmentation while preserving agency.
- **Cognitive Sovereignty:** Beyond data ownership, How do we define a formal framework to ensure AI systems enhance, rather than replace, core human reasoning?
- **Hierarchical Memory and Context Management:** Realizing fluid context will require moving beyond simple vector databases. We propose a multi-timescale memory system—immediate context (seconds), session memory (hours), and long-term knowledge (weeks)—which will be prototyped post-MVP.
- **The Serendipity Engineering Challenge:** The core technical risk of our Insight Engine is the signal-to-noise ratio. Engineering genuine insights without overwhelming the user is a long-standing challenge. Our proposed solution involves a combination of graph-based analysis and mechanisms like the "Focus Score" to manage interruptibility, alongside a feedback loop for continuous relevance training.
- **The Ubiquitous Personal OS Challenge:** As compute moves into the environment, a new OS paradigm is needed. A key research challenge is developing the architecture for a personal, sovereign OS that can travel with the user and maintain a coherent state across a multitude of devices. This includes solving the problem of peer-to-peer discovery and communication between Wise instances (a **"WiseNet"**), allowing for resilient, local-first collaboration without reliance on centralized servers.

7 The Path Forward

Our roadmap is structured to balance ambitious vision with incremental validation.

- **Near-term (0-12 months):** Complete the Wise Sovereign Edition. Execute initial user studies. Open-source the core orchestration framework.
- **Medium-term (12-24 months):** Expand to all Cognitive Lenses. Develop the initial agent SDK. Implement the persistent SQLite/LanceDB "Second Brain."
- **Long-term (24+ months):** Begin RD for the ambient, post-screen OS paradigm. Develop the "WiseNet" peer-to-peer protocol and the federated learning pipeline for the "Knowledge Refinery."

8 Conclusion

The work of visionaries like Bush and Engelbart was not about predicting the future, but about articulating a future worth building. With EngelBERT, we are building the OS to deliver it. This is not just a better way to work; it is a new way of thinking. The Wise Sovereign Edition is the first stone laid in the foundation of a cathedral for the mind—a cathedral that will ultimately transcend the screen itself. By offloading the mechanical friction of information processing while preserving human creativity and control, we envision a future where the cognitive partnership between human

and machine enables a new renaissance of thought. After 60 years of GUIs, we believe we can finally build a computer that listens—and thinks with you.

This vision paper aims to catalyze research in sovereign, proactive, and ambient cognitive computing. We invite the research community to join us in building this cathedral.

References

Aarts, E., & De Ruyter, B. (2009). New research perspectives on ambient intelligence. *Journal of Ambient Intelligence and Smart Environments*, *1*(1), 5–14.

Bush, V., et al. (1945). As we may think. *The atlantic monthly*, *176*(1), 101–108.

Chase, H. (2022). *Langchain. github.*

Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, *5*, 4–7.

Engelbart, D. (1962). Augmenting human intellect: A conceptual framework. summary report. *Stanford Research Institute, on Contract AF*, *49*(638), 1024.

Fulbright, R. (2019). Calculating cognitive augmentation–a case study. In *Augmented cognition: 13th international conference, ac 2019, held as part of the 21st hci international conference, hcii 2019, orlando, fl, usa, july 26–31, 2019, proceedings 21* (pp. 533–545).

Goldhaber, M. H. (1997). The attention economy and the net. *First Monday*.

Kahneman, D. (2011). Thinking, fast and slow. *Farrar, Straus and Giroux*.

Kay, A. C. (1996). The early history of smalltalk. In *History of programming languages—ii* (pp. 511–598).

Licklider, J. C. (2021). Man–computer symbiosis (1960).

Miehling, E., Ramamurthy, K. N., Varshney, K. R., Riemer, M., Bouneffouf, D., Richards, J. T., ... others (2025). Agentic ai needs a systems theory. *arXiv preprint arXiv:2503.00237*.