

Operating Intelligence: Transforming the Operating System into a Proactive, Human-Centered Companion

Thomas Gere¹

¹London, United Kingdom , thomas@agylgroup.com

December 12, 2025

Abstract

Operating systems have historically been reactive platforms that wait for user commands and manage resources deterministically. We introduce **Operating Intelligence (OI)**, an OS-integrated, proactive intelligence layer that anticipates needs, coordinates sensor and application ecosystems, and supports lifelong learning, health, and professional growth while preserving human agency. Motivated by a parent’s desire for a safe, proactive tutor for an eight-year-old and by workplace needs for healthier, more creative workflows, we define OI, present a reference architecture, describe interaction and user-interface principles that support fully non-visual, fully verbal, and multimodal operation, and outline representative scenarios, an evaluation methodology, ethics and governance mechanisms, and a practical roadmap for responsible deployment. We also examine OI’s role in enabling stable workforce transitions through embedded upskilling, role redesign, and lifestyle reorganization, and integrate AGI alignment and safety considerations into the architecture and governance model.

1 Introduction

Traditional operating systems (OS) are reactive: they execute user commands, schedule processes, and manage hardware resources. Contemporary life, however, is continuous and context rich. Wearables, environmental sensors, and always-connected services produce streams of signals that, if responsibly harnessed, enable anticipatory support. We propose **Operating Intelligence (OI)** as a principled transformation of the OS into an attentive partner that anticipates needs, stages resources, and scaffolds human goals—learning, well-being, creativity, and collaboration—while enforcing explicit, revocable permissions and transparent explanations.

A motivating vignette clarifies the stakes. A parent wishes for a computer that acts as a tutor and advisor for an eight-year-old: one that adapts lessons to the child’s pace, nudges healthy study habits, and prevents exposure to harmful content without undermining curiosity. The same companion layer should extend into adulthood and work, helping professionals prioritize, sustain healthy routines, pursue targeted training, assure quality, and stimulate imagination. The central research question is how an OS-level intelligence layer can deliver anticipatory, personalized assistance across life stages while guaranteeing privacy, explainability, and equitable access.

This manuscript articulates a definition and design commitments for OI, describes a reference architecture and interaction model that supports fully non-visual and fully verbal operation as well as multimodal experiences, presents representative scenarios and an evaluation plan, and outlines governance and deployment pathways. It also examines OI’s potential to ease socioeconomic

transitions caused by automation by embedding upskilling and role redesign into everyday workflows and integrates alignment and AGI-relevant safety practices into the architecture.

2 Background and related work

Operating systems research established the substrate for isolation, scheduling, and device control. Ubiquitous computing and ambient intelligence proposed systems that fit human rhythms and reduce cognitive friction. Human-computer interaction (HCI) developed methods for attention-sensitive design and accessibility. Agent systems and modern assistants introduced new forms of proactivity and delegation. Advances in on-device models, federated learning, multimodal interfaces, and low-latency networks make proactive, privacy-preserving assistance feasible at scale.

Despite these advances, existing approaches are often fragmented: assistants are siloed in applications, safety policies are bolted on, and personalization can overreach or erode trust. Operating Intelligence positions intelligence and governance within the OS substrate so anticipatory behaviors are first-class, auditable, and consistent across contexts. Forecasts and industry roadmaps for increasingly capable agents motivate designing OI with both near-term safety features and extensible governance for higher-capability agents; incorporating alignment research, deployment practices, and policy guidance helps ensure that OI remains robust under a range of capability trajectories.

3 Definition and design commitments

Operating Intelligence is an OS-integrated intelligence layer that proactively orchestrates resources, interactions, and protections to optimize for human goals—learning, well-being, creativity, and collaboration—subject to explicit, revocable permissions and transparent rationales.

Design commitments:

- **Proactivity with permission.** Anticipate needs and propose actions only within user-defined scopes and time bounds.
- **Human agency first.** Assist and scaffold; do not coerce or replace human judgment.
- **Privacy by design.** Default to local processing, minimize data collection, and provide auditable logs.
- **Explainability.** Provide plain-language rationales, confidence indicators, and easy undo.
- **Safety and alignment.** Enforce age-aware safeguards, adversarial testing, and red-teaming for high-risk behaviors.
- **Equity and accessibility.** Ensure multimodal parity, localization, and low-bandwidth/energy-efficient modes.

4 Reference architecture

The OI architecture interleaves intelligence and governance with core OS services and is organized into six interacting layers: hardware and sensing, trusted runtime, intelligence kernel (OI core), context fusion, interaction, and governance.

4.1 Layers

Hardware and sensing. Interfaces with CPUs/NPUs, secure enclaves, wearables, environmental sensors, and radios (Wi-Fi, Bluetooth, 5G). Hardware identity and attestation support trust.

Trusted runtime. Provides isolation, scheduling, secure I/O, and storage primitives that the intelligence kernel uses to enforce policies and sandboxed actions.

Intelligence kernel. Contains a policy engine, contextual user models, goal and constraint representations, multi-agent orchestration, and explainability services. It subscribes to permissioned context streams, generates candidate plans with confidence scores and rationales, and submits plans to the governance layer for enforcement.

Context fusion. Local-first fusion combines signals from wearables, environmental sensors, and application telemetry. Federated learning and differential privacy are available for cross-device model improvement under explicit consent.

Interaction layer. Multimodal interfaces (visual, voice, haptic, AR/VR) present proposals, rationales, and controls. UI primitives include the *proposal card*, *confidence badge*, *undo token*, and *explainable trace*.

Governance layer. Consent management, transparency logs, safety constraints, escalation pathways, and certification for third-party agents live here. All proactive actions are logged with human-readable rationales and machine traces.

4.2 Data flow and capability gating

The intelligence kernel proposes actions; the governance layer enforces constraints and logs decisions. Low-confidence proposals default to suggestions requiring explicit approval; high-confidence safety interventions may be automatic but are immediately explained and logged. When connectivity or confidence is insufficient, OI falls back to passive assistance and surfaces options to the user.

Capability gating and staged escalation. Architecturally, OI must enforce capability gating: a staged permission model that requires progressively stronger attestations, human approvals, and governance checks before enabling high-impact autonomous actions. Capability gating includes conservative defaults, staged sandboxing, attestation of model provenance and version, and explicit escalation paths that require human sign-off for actions above defined impact thresholds. This pattern reduces the risk of inadvertent high-impact behavior as agent capabilities increase and supports certification and audit workflows recommended by contemporary safety research and policy.

5 Interaction model and user interface

The interaction layer is the primary locus where proactivity, consent, and explainability meet human perception and action. OI’s UI design ensures modality parity, progressive disclosure, and reversible actions so proactivity empowers rather than surprises.

5.1 Modalities

Fully non-visual operation. Visual elements map to semantic labels and structured audio summaries. Voice-first conversational dialogs enable multi-turn negotiation; haptic patterns convey status and transitions; programmable gestures and voice macros provide discoverable shortcuts.

Fully verbal operation. OI can initiate spoken suggestions with concise rationales and simple accept/decline vocabularies. Dialogic negotiation allows users to refine proposals by voice. Verbal prompts respect ambient privacy and defer or downgrade when others are present.

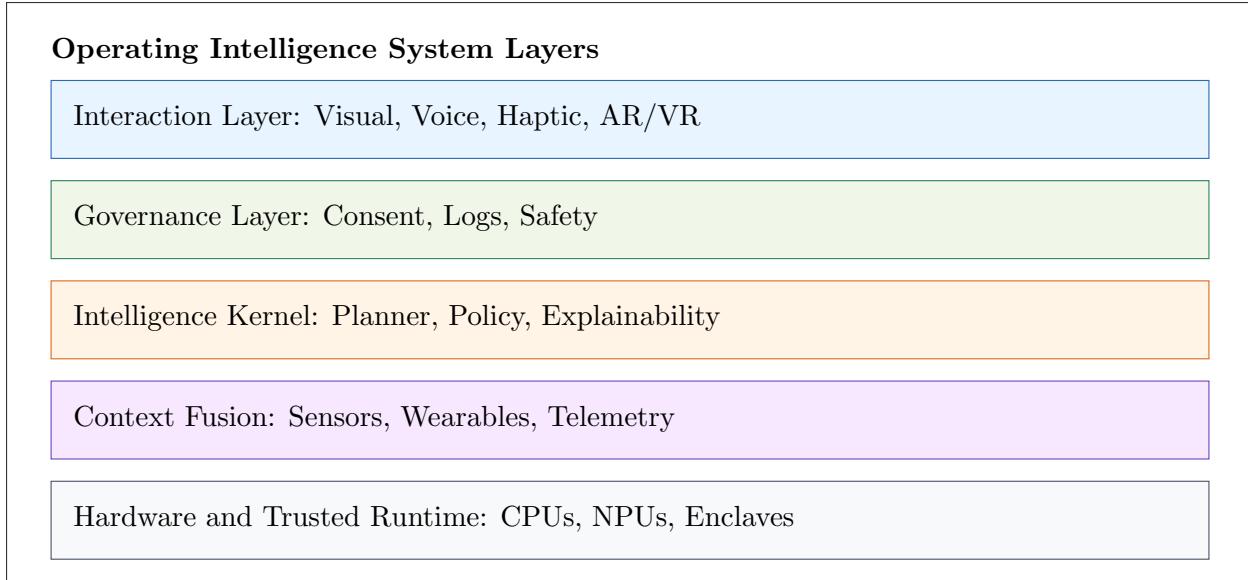


Figure 1: System layers for Operating Intelligence.

Visual and app-centric operation. OI stages applications for human input: preloading modules, opening editors with drafts, or configuring collaboration boards. Proposal cards summarize planned actions with a confidence badge and a single accept control. Overlays are non-modal by default. AR/VR interfaces use spatial annotations and ephemeral anchors for immersive training and collaboration.

5.2 Human-in-the-loop flows

For proposals that would enable high-capability or high-impact agent actions, the UI enforces multi-step human-in-the-loop flows: a concise rationale and confidence score, an evidence summary and explainable trace, an explicit approval step with role-based sign-off, and a time-bounded undo and audit record. These flows preserve agency and provide clear accountability when OI coordinates or delegates actions to powerful agents.

6 Applications and scenarios

OI's value is best illustrated through scenarios that span life stages and social contexts.

6.1 Child tutor and protector

OI curates lessons, adapts difficulty, interleaves play and reflection, and enforces age-appropriate content policies. Guardians receive privacy-preserving summaries and can adjust boundaries without accessing raw child data. Metrics: learning gains, time-on-task quality, safety incidents, guardian satisfaction.

6.2 Neurodiverse learner

A teenager with sensory sensitivity configures OI to reduce notification frequency and prefer predictable routines. OI schedules study blocks with consistent cues and provides explicit transition

Interaction Modalities and UI Primitives

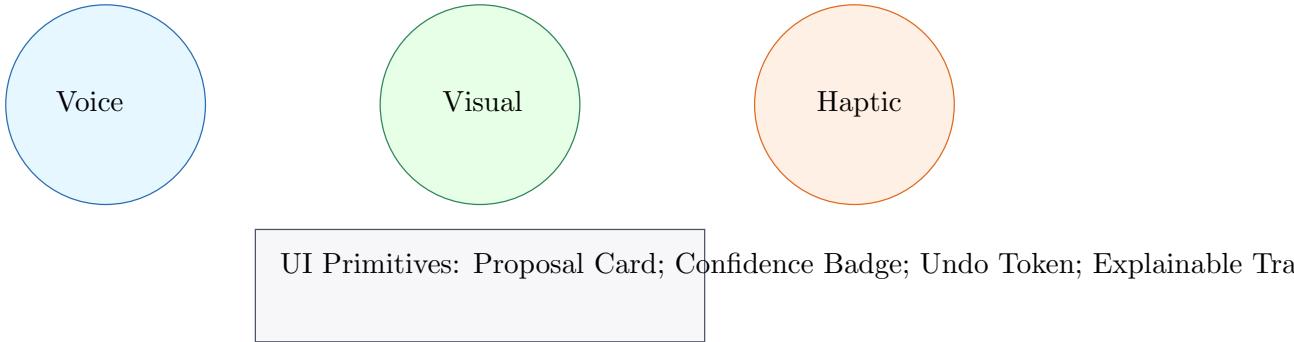


Figure 2: Interaction modalities and UI primitives.

Scenario Timeline Example Child Tutor



Figure 3: Scenario timeline (child tutor example).

prompts. Metrics: task completion, stress indicators, override frequency.

6.3 Work companion

A mid-career engineer receives a prioritized task list that balances deadlines and cognitive load. OI schedules deep-work blocks, suggests micro-learning modules in AR for a new IoT API, and performs pre-flight checks on a release candidate. Metrics: productivity (throughput and quality), health adherence (breaks), training completion, perceived autonomy.

6.4 Low-bandwidth household

OI operates in offline-first mode with energy-efficient inference and compact summaries. It synchronizes only essential model updates under explicit consent. Metrics: equitable performance, user satisfaction under constrained connectivity.

7 Socioeconomic impact and workforce transition

Automation and agentic systems will reshape the social and work fabric. OI can act as an enabling layer for stable, humane transitions by coordinating upskilling, redesigning roles, and supporting

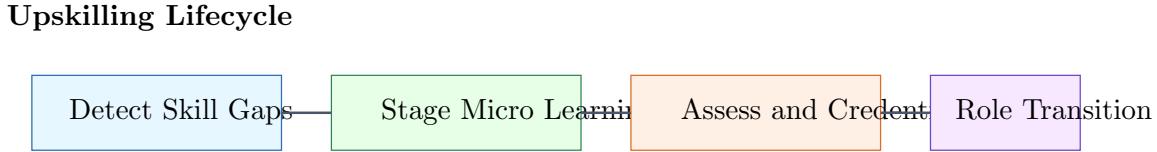


Figure 4: Upskilling lifecycle: detection, staging, credentialing, role transition.

lifestyle reorganization.

7.1 Embedded upskilling

OI detects skill gaps (with consent) by observing task patterns and outcomes and maps them to modular learning pathways that fit the user’s schedule and cognitive load. It stages just-in-time learning—short interactive modules, AR/VR simulations, or guided practice—immediately before or after relevant tasks. OI captures privacy-preserving evidence of mastery and suggests micro-credentials or project opportunities that demonstrate new capabilities.

7.2 Role redesign and task reallocation

OI analyzes workflows to identify automatable, augmentable, and human-centric tasks. It proposes redesigned role descriptions that emphasize supervision, creativity, and social coordination while delegating routine execution to agents. Human-in-the-loop orchestration reduces cognitive load by summarizing agent activity, highlighting anomalies, and providing concise decision aids so humans can supervise multiple agents effectively.

7.3 AGI-accelerated transition scenarios

Rapid capability advances could accelerate automation of complex tasks and shorten transition windows for affected workers. OI must therefore support contingency workflows: accelerated micro-learning tracks, prioritized credential portability, temporary schedule and income planning tools, and rapid employer-sponsored retraining pathways. Organizational integrations should include privacy-preserving diagnostics that identify emergent skill gaps, and public-sector interfaces that connect displaced workers to accredited training and social supports.

8 Ethics, privacy, governance, and AGI alignment

Proactivity demands robust ethics. OI operationalizes consent through granular, time-bounded permissions and local-first defaults. Explainability is mandatory: every intervention includes a plain-language rationale and an undo path. Safety boundaries are age-aware and topic-sensitive,

with escalation to human support rather than automated replacement. Fairness requires regular audits for disparate impact and participatory design with diverse stakeholders. Accountability is enforced through tamper-evident logs, independent audits, and clear vendor and organizational responsibilities.

8.1 AGI alignment, verification, and guardrails

As agent capabilities advance, OI’s governance primitives must incorporate alignment and verification techniques from AGI safety research. Practically, this means integrating interpretability tools (feature and activation inspection, concept attribution), formal verification where feasible (safety properties for constrained modules), adversarial red-teaming pipelines, and capability gating. OI should require provenance and attestation for any high-capability model (model lineage, training data summary, evaluation artifacts) and maintain tamper-evident logs of model decisions and governance checks. Where full formal guarantees are infeasible, OI relies on layered defenses—conservative defaults, human oversight, adversarial testing, and continuous monitoring—to reduce risk. These mechanisms enable independent audits and support regulatory compliance.

9 Evaluation methodology and AGI-specific testing

Assessing OI involves user outcomes, human factors, system performance, privacy and safety, and equity. We recommend mixed methods: controlled lab studies for interaction and cognitive load; longitudinal field trials (8–12 weeks) in education and workplace pilots to measure learning outcomes, well-being indicators, and productivity; human factors evaluations for trust calibration and perceived agency; system metrics for latency, reliability, and energy; privacy and safety red-teaming; and equity audits across demographics and connectivity conditions.

9.1 AGI-specific evaluation

AGI-specific evaluation augments these methods with adversarial testing, interpretability audits, and verification protocols. Pre-deployment checks include model provenance and attestation, safety property checklists, and baseline interpretability analyses. Red-teaming pipelines combine automated adversarial generation with human expert exercises to probe for unsafe behaviors. Human-in-the-loop verification enforces multi-role sign-off for high-impact proposals and requires explainable traces that human reviewers can inspect quickly. Continuous monitoring uses compact, privacy-preserving telemetry and anomaly detectors to flag deviations. Tamper-evident logs and incident playbooks support forensic analysis and independent audits.

10 Roadmap and implementation sketch

A practical path to OI unfolds in phases. Phase 1 introduces context-aware recommendations, safety features, and accessibility baselines. Phase 2 refactors the substrate to host an intelligence kernel with on-device models and explainability. Phase 3 realizes full OI with rich sensor fusion, AR/VR interfaces, and edge/5G integration under comprehensive governance and capability gating. Phase 4 scales the ecosystem with open APIs, certified third-party agents, and interoperability standards. Public-interest pilots in education, well-being, and workplace contexts anchor development in real needs and diverse populations.

A minimal prototype includes an on-device policy engine, a context fusion module that ingests wearable and environment signals, a planner that proposes actions with rationales, and a consent UI.

Developer contracts and UX primitives (proposal card, confidence badge, undo token, explainable trace) enable third-party integration while preserving safety.

11 Conclusion

Operating Intelligence reframes the operating system as a proactive, human-centered companion that anticipates needs, respects boundaries, and elevates learning, well-being, creativity, and collaboration. Built on automation, immersive interfaces, sensor fusion, and modern connectivity, and governed by consent, privacy, and inclusion, OI offers a credible path to computers that adapt to people across ages and social contexts. Its success depends on interdisciplinary collaboration, rigorous evaluation, and adaptive governance that incorporates alignment and verification practices as agent capabilities advance. By committing to ethical design and equitable deployment, we can shape a computing paradigm that is both powerful and humane.

Acknowledgments

Thanking support tools such as Copilot in Think deeper mode, which helped me articulate my ideas and research further.

References

- [1] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.
- [2] Yuntao Bai et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.
- [3] Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.
- [4] European Commission. Proposal for a regulation laying down harmonised rules on artificial intelligence, 2021.
- [5] Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*. Foundations and Trends in Theoretical Computer Science, 2014.
- [6] Ray Kurzweil. *The Singularity Is Near: When Humans Transcend Biology*. Viking, 2005.
- [7] OECD. Oecd ai principles. 2019.
- [8] Andrew S. Tanenbaum. *Modern Operating Systems*. Pearson, 2014.
- [9] Mark Weiser. The computer for the 21st century. *Scientific American*, 265:94–104, 1991.
- [10] Daniel Weitzner et al. Privacy and security by design. *Communications of the ACM*, 2016.