

The Next Frontier of AI-Driven Client/Customer Experience: The Human/AI Interaction

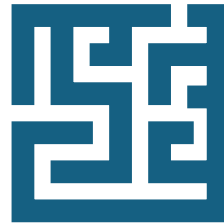


Besa Bauta, Ph.D., MPH, MSW
CIO | JBFCS

Problem + Experience Gap + Promise



The Reality: Most AI applications today are **reactive**. They answer FAQs, routes tickets, and follow scripted flows. It's *service but not experience*.



The Challenge: The moment an experience gets **complex**, frustrating, or emotional, the AI applications fail, and the human agent is left to "clean up."

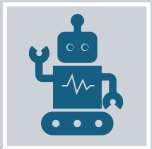


The Promise: Moving toward a future where AI and humans **collaborate** to anticipate needs, elevate decisions, and build lasting trust.

Collaborative Frontier



Augmentation over Automation: Collaborative AI moves beyond simply replacing human effort to actively boosting human performance, especially in complex tasks requiring creativity, strategic thinking, and emotional intelligence.



Context and Intent: New generation of AI systems need to understand the *why* behind a request (intent, emotion, and situational context) rather than just the *what* (the literal words).



The Partnership: Shifting relationships from a master/tool dynamic to a **human-AI partnership** where both entities contribute their complementary strengths to achieve shared goals.

The Next Frontier: Human-AI Collaboration

Reactive AI (Today's Standard)

Answers questions

Follows scripts

Automates tasks

Saves cost

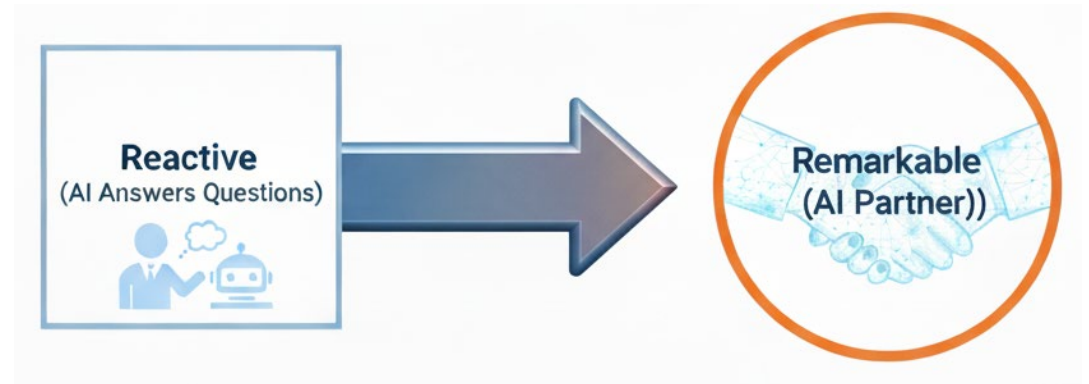
Collaborative AI (The Frontier)

Anticipates needs

Adapts to context & emotion

Enhances human judgment

Creates value & connection



- The future isn't *AI supporting* users; it's *AI partnering* with them.
- AI's new role is to **reduce cognitive load** for both the client and the agent, allowing them to focus on high-value connection.

Thinking and Reasoning Architectures

Human Cognitive/Brain Architecture:

- How we learn, think, and solve problems
- A natural information processing system that generates various procedures designed to reduce cognitive load and facilitate the acquisition of biologically secondary knowledge held in long-term memory.

Human Knowledge:

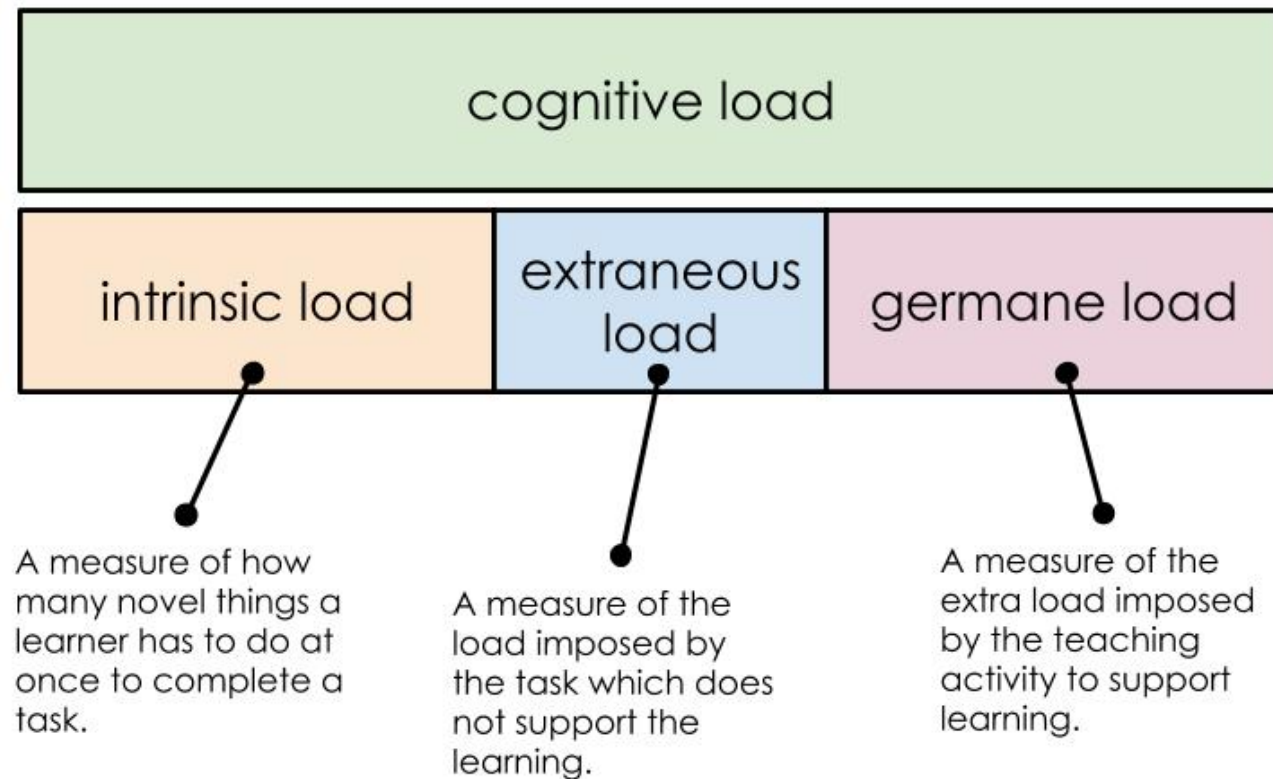
- **Biologically Primary**
 - Low cognitive load: evolved to acquire such information automatically, it does not need to be taught
 - General problem-solving, thinking skills such as learning to speak and listen in a native language,
 - Generalizing, transferring, and performing simple social skills like recognizing faces.
- **Biologically Secondary**
 - Explicitly taught and learned
 - **Higher cognitive load:** requires conscious and mental effort

Cydonia region of Mars

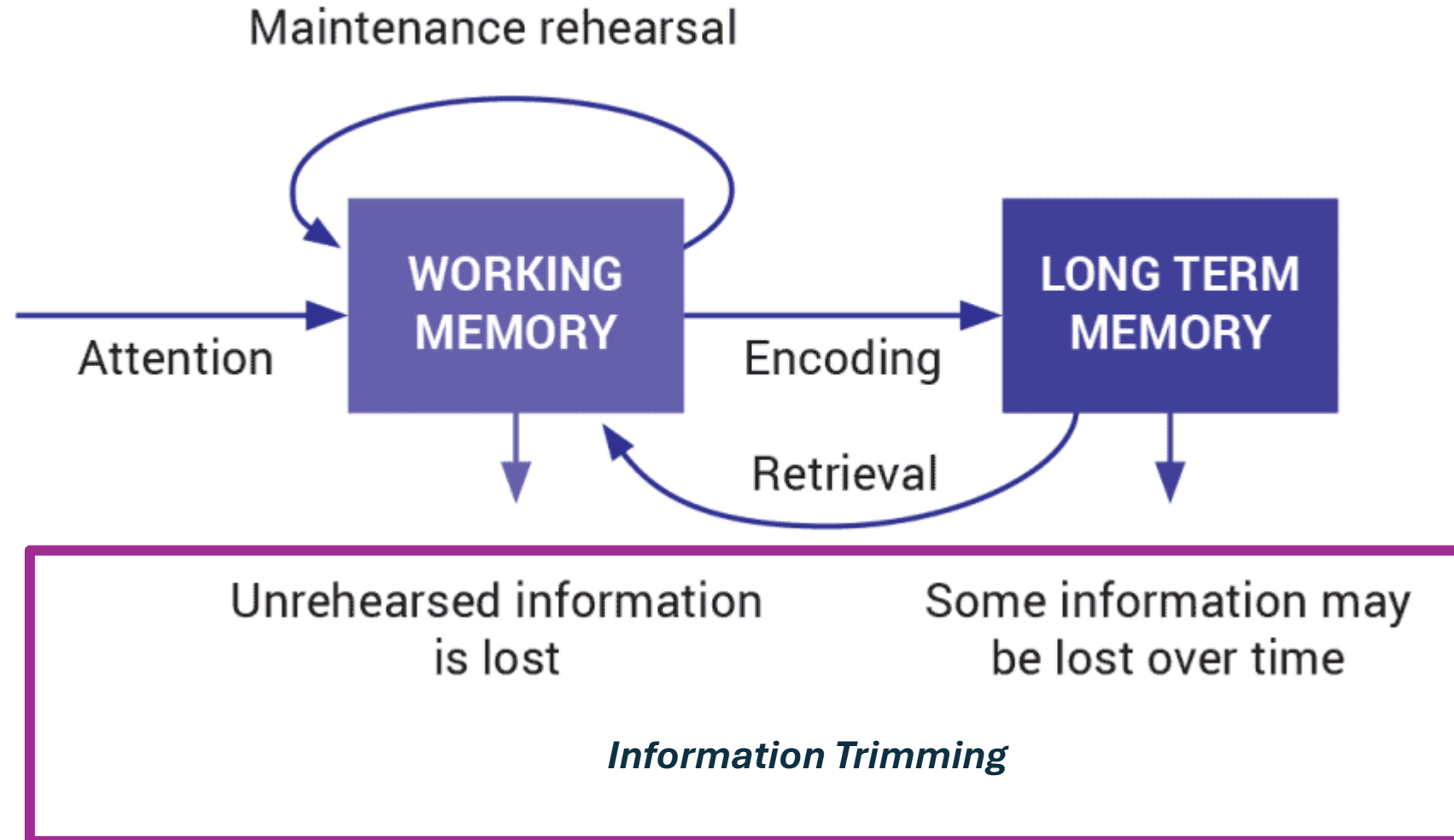


Cognitive Load Theory (CLT)

- CLT model describes memory as having three main parts: **sensory, working, and long-term**.
- Cognitive load = amount of information our working memory can process at any given time.

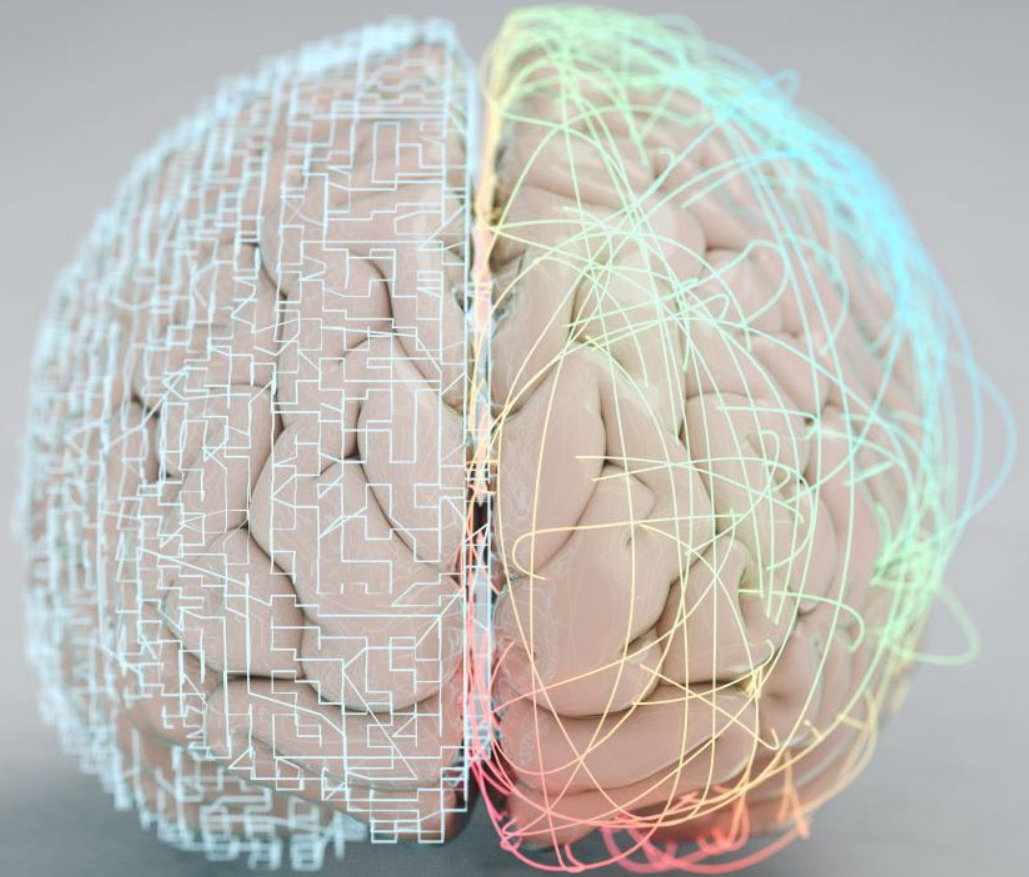


Human memory



Anthropomorphic AI

- AI is fundamentally inspired by the structure and processes of the biological brain.
- Physical components and complexity differ; the core organizational and functional features are the same.



Neurons and Nodes

The key parallels between AI architecture (specifically **Artificial Neural Networks (ANNs)**) and the brain lie in their fundamental organizational units and how they process information:

Processing Units (Neurons/Nodes):

- **Biological Brain:**
 - Billions of **neurons** (nerve cells).
- **AI Architecture:**
 - Consists of **nodes** or "**artificial neurons**" (mathematical functions) arranged in layers.
 - Both the biological neuron and the artificial node receive multiple inputs, process them, and then produce an output signal.

Connections and Weights (Synapses/Weights):

- **Biological Brain:**
 - Neurons connect at **synapses**, which determine the strength and efficiency of signal transmission using chemical and electrical signals.
- **AI Architecture:**
 - Nodes are linked by **connections**; each assigned a numerical **weight**.
 - This weight determines the influence of one node's output on the next node's input.
 - **Weights** in an ANN are the computational equivalent of the **synaptic strength** in the brain and is adjusted during learning to improve performance.

Neurons and Nodes

Layered and Parallel Processing:

- **Biological Brain:** Information is processed hierarchically across different brain regions (e.g., visual cortex, prefrontal cortex) that **operate in parallel** to handle various aspects of a task.
- **AI Architecture:** ANNs, especially Deep Learning models CNN/Convolutional Neural Networks, use multiple *hidden layers* (*deep layers*) to process data.
- Information flows through these layers in a complex, **parallel fashion**, with each layer extracting increasingly abstract features.

Learning and Adaptation (Plasticity/Training):

- **Biological Brain: Neural plasticity** is the ability of synapses to strengthen or weaken over time, allowing the brain to learn, form memories, and adapt to new experiences.
- **AI Architecture:** The network **learns** by adjusting the **weights and biases** of its connections through training algorithms (like backpropagation) to minimize errors.
- This adjustment is directly analogous to biological learning.

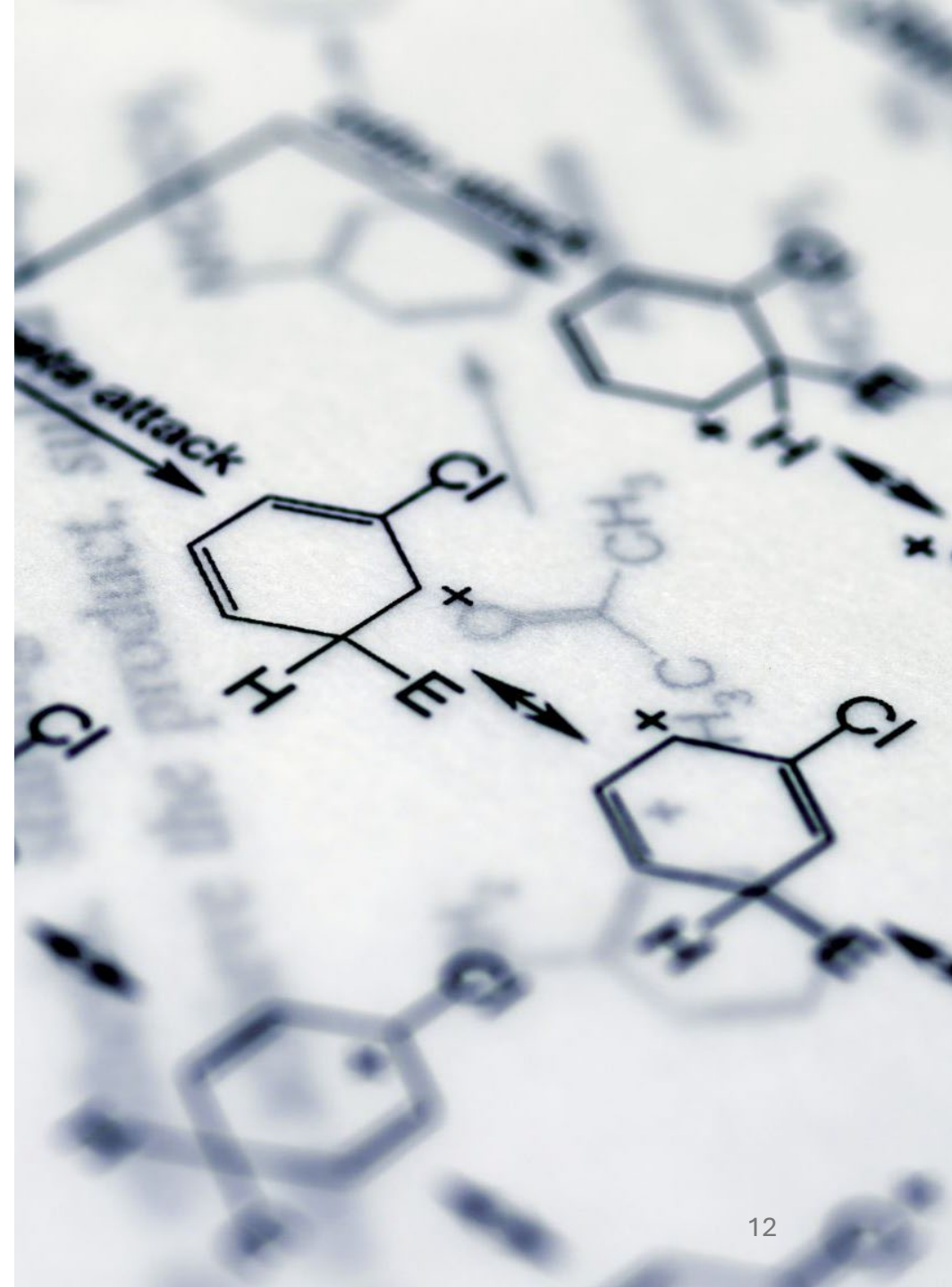
Design Differences

Despite the inspiration, the implementation in a machine differs significantly from biology:

Aspect	Biological Brain	Artificial Neural Networks (ANNs)
Physical Basis	Wetware (Biological cells, chemical and electrical signals)	Software (Algorithms, data structures, and mathematical functions running on silicon chips)
Signal Transmission	Spikes (discrete electrical pulses, often asynchronous).	Continuous values (analog-like, usually synchronous or batch-processed).
Scale	~86 billion neurons, ~100 trillion synapses.	Up to a few trillion parameters (weights) in the largest models.
Power Consumption	Extremely energy efficient (runs on ~20 watts).	Requires massive computational power and energy.
Learning	Mostly one-shot learning and continuous, life-long adaptation.	Requires vast amounts of data and many iterations (training epochs).

Analog vs Digital

- **Brain:** Operates with **analog**, continuous biochemical processes that are "just reliable enough." Biological neurons have a huge variety of non-linear responses.
- **AI:** Operates on **digital**, binary logic (0s and 1s) and high-precision floating-point numbers, **requiring more energy** for precise, deterministic operations.



Powering Brains: The Efficiency Gap

Aspect	Human Brain	AI Machine (Modern Systems)
Continuous Power Draw (Inference)	~12-20 Watts (Less than a dim light bulb).	Tens of thousands of Watts (for large server racks/GPUs).
Energy for Training (LLM like GPT-3)	NA(Continuous learning from food energy)	~1,300 Megawatt-hours (MWh) (Equivalent to the annual consumption of ~130 US homes).
Computational Efficiency	Extremely high (performing an Exaflop equivalent with minimal power).	Millions of times less efficient than the brain on many complex tasks.

Key Differentiators

Human Brain:

- Marvel of evolutionary optimization.
- 2% of adult body mass consumes 20% of body's total energy (highly efficient power usage).
- 20 Watts = $\approx 10^{18}$ [one billion operations per second]

Artificial Neural Networks (ANNs):

- Run on silicon chips (e.g., CPU, GPUs, NPUs, TPUs)
- Based on von Neumann Architecture which separates processing unit from memory
- Requires enormous energy to move data back and forth – key source of inefficiency

Anthropomorphic AI: Understanding the Human Element



How AI is Evolving:



Sentiment Analysis to Emotional Context: Not just *what* is said, but *how* it's said (tone, pace, hesitation) to infer emotional state.



"Theory of Mind" Capabilities: AI models that can infer a client's beliefs, goals, and knowledge, the *why* behind the *what*.



Example: A client mentions a flight delay; the AI understands their *goal* (making a connecting meeting) and proactively offers a local taxi voucher.



Actionable: Design AI to listen for signal, not just keywords.

The Art of Decision Augmentation

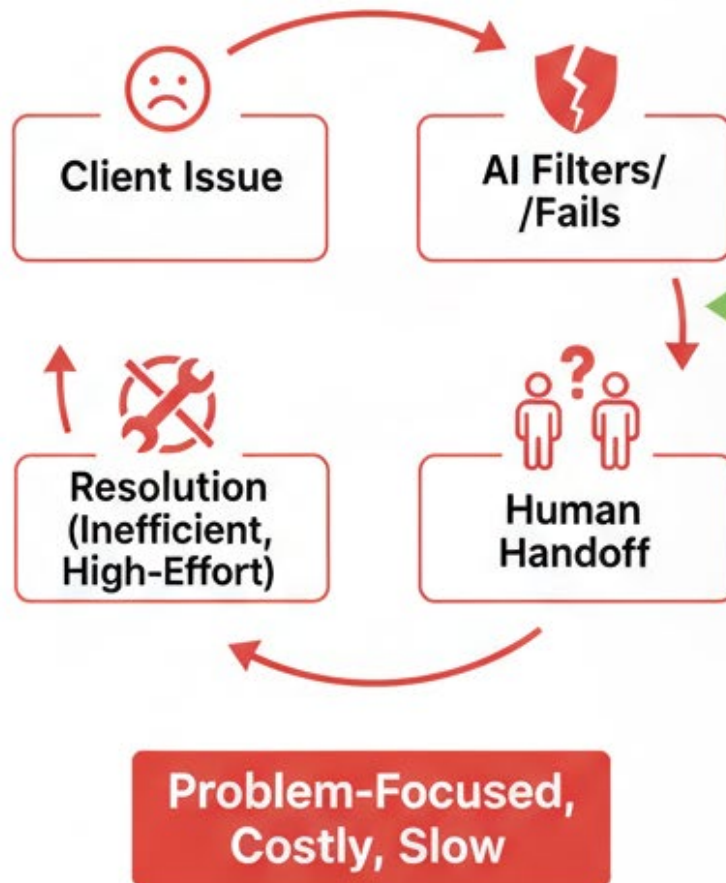
Using AI to **reduce cognitive load** and empower better human judgment.

Key Mechanisms:

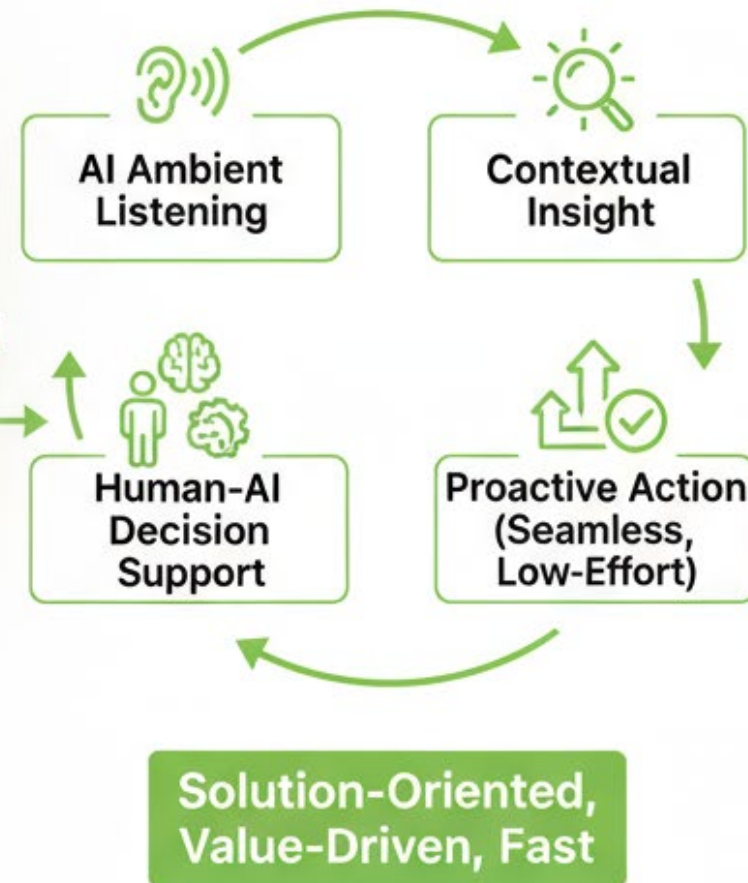
- Ambient Listening: AI passively analyzes ongoing conversations (voice, chat) in real-time.
- Conversational AI for Agents: Provides next-best-action prompts, relevant data, and compliance checks *as the conversation happens*.
- Results: The human agent spends less time searching for data and more time connecting and applying judgment to the complex nuances of the client's request.

Anthropomorphic AI: 2.0

REACTIVE CYCLE



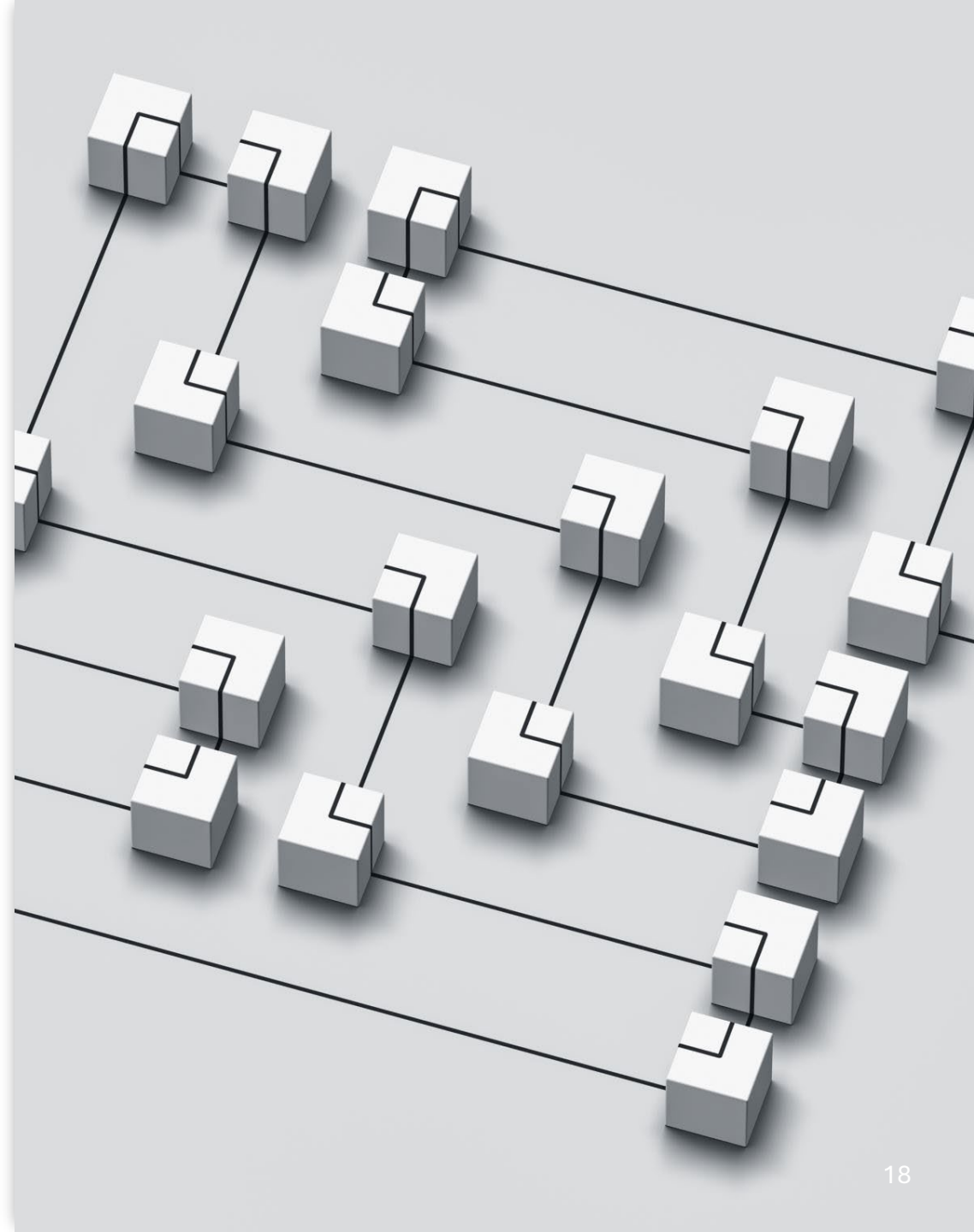
REMARKABLE LOOP



Trust Through Transparency

The Trust Mandate:

- **Show Your Work:** Recommendations and decisions shouldn't be a black box. Providing reasoning and data sources behind suggestions.
- **Relevance & Alignment:** models need an internal 'values' framework to ensure prioritization of outcomes that align with client well-being and organizational ethics.
- **The Usability Loop:** Make it easy for agents to accept, reject, or modify AI suggestions and ensure that feedback is incorporated instantly to improve the model.
- AI must be interpretable and aligned with human values to be adopted.



From Today to Tomorrow



Action 1: Redefining Development Goals: Stop automating services; start *augmenting human capability*. Focus AI on the most complex, emotional, and high-value interactions.



Action 2: Investing in Context: Treating data context (not just data volume) as most valuable assets. Prioritizing technologies that mirror human emotion-sensing capabilities.



Action 3: Establishing a Trust Framework: Demanding interpretable/integrated AI infrastructures and systems. Ensuring Systems can explain their recommendations and align with human goals and ethical values.

The Ultimate AI-Driven Experience

Remarkable isn't about eliminating the human touch.

It's about empowering it with the right information, timing, and profound empathy.





Thank you!

Besa.Bauta@nyu.edu
or bbauta@jbfcs.org