CHAPTER 1

THE QUESTION BEFORE THE MACHINE

Can artificial intelligence form an opinion of its own?

The first sign of an engineered mind is not the answer it gives, but the way it weighs what it does not know.

Ibrahim Youssef Noshokaty

The Magic Wand and the Missing Mind

Every age chooses an object onto which it projects its greatest hopes and its deepest fears. For the twenty-first century, that object is artificial intelligence. We ask it to write, design, diagnose, calculate, teach, translate, compose, predict, and protect. To some, it is a magic wand. To others, it is a rival waiting behind the curtain. Between fascination and anxiety, however, a more important question is often lost: when a machine produces an answer, has it understood anything - or has it merely arranged the traces of what we already know?

This book was born from that question. It is not a declaration that present-day machines are conscious, nor a promise that a particular software system has crossed the border into subjective experience. Contemporary consciousness science does not justify such certainty. Researchers still disagree about the necessary and sufficient conditions for consciousness, and recent work therefore recommends theory-based indicators rather than theatrical tests based only on fluent behavior.[7,8] The ambition of this book is more disciplined and, in my view, more exciting: to investigate the engineering conditions that may allow an artificial system to integrate sensation, uncertainty, memory, values, prediction, and self-correction into something more coherent than a retrieved answer.

Since 2015, my work on the ENOSH project has moved through cinema, narrative generation, acoustics, signal analysis, and models of human cognition. Shakespearean drama offered a demanding training ground because its language contains ambiguity, motive, conflict, silence, reversal, and emotion. Cinema added another dimension: an intelligent system did not merely need sentences; it needed scenes, timing, visual continuity, character intention, and consequences. The experiments were imperfect, but they changed the direction of my research. I began by asking whether a machine could write. I ended by asking what must exist inside a machine before its writing can be called a decision rather than an echo.

That distinction - between output and opinion - is the doorway to this book.

Creativity Is an Event, Not a Permanent Condition

Human creativity is frequently romanticized as a mysterious flame. Yet creative thought is neither a single chemical event nor the activity of one isolated brain center. Neuroimaging research increasingly describes it as a changing coordination among large-scale networks involved in spontaneous idea generation, relevance detection, and executive control.[6] A person does not remain continuously creative. Attention shifts. Energy changes. Memory reorganizes. An idea appears, is tested, rejected, reshaped, and sometimes returns in another form.

This leads to the first provocative question: can artificial intelligence remain in a state of creativity? The answer depends on what we mean by creativity. If creativity is only the production of something statistically unusual, machines can already surprise us. If it is the useful recombination of known elements, they can do more. But if creativity also requires a situated intention - an awareness of why a new form matters, whom it may affect, what it risks, and whether it should be pursued - then novelty alone is not enough.

Shakespeare did not invent every word he used; he transformed human conflict into structures that still disturb us. Christopher Nolan does not create cinema from emptiness; he reorganizes time, memory, and causality. Maxwell's demon was not a physical creature but a thought experiment that forced science to examine information and entropy. Ahmed Zewail did not create time; he opened a window onto chemical events at the femtosecond scale. In every case, creation emerged from a meeting between inherited knowledge and a new frame of meaning. A machine may generate combinations without fatigue, but an engineered mind must also evaluate significance.

I therefore use creativity in this book as a three-part process: generation of alternatives, evaluation of their value and consequences, and revision in light of uncertainty. The first part may be computationally abundant. The second and third are where the problem of an artificial opinion begins.

From Information to Opinion

A simple proverb exposed the weakness of my early assumptions. I asked more than one conversational system to explain the expression, “A cat has nine lives.” The answers were fluent, similar, and apparently confident. Their agreement initially looked like evidence of understanding. A deeper search suggested something less impressive: the systems were converging on familiar explanations that were already highly visible in their information environment. They had found an answer. They had not demonstrated that they could weigh competing meanings, detect the cultural frame, or explain why one interpretation should be preferred over another.

Agreement is not judgment. Fluency is not understanding. Retrieval is not opinion.

The lesson was not that databases are useless. Human beings also depend on inherited language, stories, teachers, and archives. The lesson was that information must pass through a process of evaluation before it becomes a responsible opinion. A useful artificial opinion should be provisional, explainable, sensitive to context, explicit about uncertainty, and open to revision. It should distinguish between evidence and popularity. It should be able to say, “This is the most likely interpretation, these are the alternatives, this is what I do not know, and this is why the distinction matters.”

For this book, I define an artificial opinion as a context-bound decision state produced by five operations: integrating evidence, assigning degrees of confidence, applying domain constraints,

estimating consequences, and preserving the possibility of revision. This definition does not prove subjective consciousness. It describes an engineering threshold beyond simple retrieval: a system that can organize uncertainty into an accountable position without pretending that uncertainty has disappeared.

Fuzzy Logic: A Language for the In-Between

Classical logic is powerful because it draws clean boundaries: true or false, member or non-member, one or zero. Human situations often refuse to cooperate. A sound may be somewhat urgent. A patient may be moderately unstable. A statement may be culturally plausible but weakly supported. A face may appear familiar without being identifiable. Lotfi A. Zadeh's theory of fuzzy sets gave mathematics a rigorous language for graded membership between zero and one.[1] Mamdani-type controllers later showed how linguistic rules could be translated into practical control behavior,[2] while Takagi-Sugeno models provided another influential way to represent nonlinear systems using fuzzy rules and local mathematical relations.[3]

Fuzzy logic is not simply probability. Probability asks how likely an event is; fuzzy membership asks to what degree a description applies. The difference matters. There may be a 70 percent probability of rain tomorrow, but today's air may belong to the category “humid” with a membership value of 0.70. One statement concerns uncertainty about an event. The other concerns the graded fit between an observation and a concept.

In the ENOSH architecture, fuzzy reasoning is not presented as consciousness itself. It is a candidate mechanism for the preconditions of opinion formation. It allows a system to represent “partly true,” “insufficiently known,” “highly urgent,” or “morally conflicting” without crushing these states into premature yes-or-no answers. Defuzzification can convert the resulting fuzzy state into an actionable value when action is required. But that conversion does not create a conscience. It merely makes a graded decision executable.

This is where the second edition draws a boundary that the first edition left too vague: uncertainty processing, creativity, emotion recognition, and self-monitoring may contribute to an architecture of machine awareness, but none of them alone demonstrates subjective experience. The design problem is to assemble them transparently, test each function independently, and refuse to call the whole system conscious merely because it speaks as if it were.

Can Sensory Experience Become Data?

My own path into this problem begins with sound. As an electro-acoustics researcher, I have spent much of my life watching physical vibration become information. A pressure wave reaches the outer and middle ear, sets the cochlear mechanics in motion, and is transduced by hair cells into neural activity. From there, auditory pathways analyze timing, frequency, intensity, spatial cues, patterns, and meaning. The cochlea is not “inside the brain,” and the journey is not a simple translation of letters. The auditory system receives no written word. It receives structured energy over time.

Consider the word “what.” Its dictionary form is only the beginning. Spoken quickly, slowly, sharply, softly, with a rising pitch or a lengthened vowel, it can signal a request, disbelief, irritation, fear, refusal, or humor. Prosody - the pattern of pitch, rhythm, stress, intensity, and duration - carries information about emotion and intention that the transcript alone may lose. Neuroscience describes vocal-emotion processing as a multistage interaction among sensory, cognitive, and affective systems rather than a single-channel decoding act.[5] This is also the territory that affective computing has explored for decades: systems that recognize, model, or respond to human affect without assuming that recognition is identical to feeling.[4]

That distinction is essential. A microphone can measure pressure. A model can estimate pitch, speech rate, spectral tilt, or emotional category. A camera can classify an expression. A physiological sensor can detect a pattern. None of these measurements proves that the machine hears, sees, suffers, or enjoys as a human does. Sensor data can support a richer model of the world; it cannot, by itself, certify an inner world.

The engineering opportunity is nevertheless enormous. Instead of converting speech immediately into text and discarding the acoustic evidence, a system can preserve parallel streams: words, prosody, speaker identity, room context, uncertainty, and prior interaction. Each stream can be processed independently and then integrated. The result is not human hearing. It is a more responsible route from sound to intended meaning.



Figure 1. A proposed signal-to-meaning pathway. The system preserves acoustic evidence, context, uncertainty, and human authority instead of reducing speech immediately to text.

The Hospital Test: Intelligence Under Pressure

The weakness of a decision system becomes clearest under pressure. Imagine a hospital receiving fifteen critically injured patients at once. Fifteen operating rooms are available, but medicines, sterile instruments, transport routes, specialist teams, and preparation times are not equally distributed. A conveyor or robotic delivery route reaches the first room in one minute and the fifteenth in fifteen. The naive system follows the order of the rooms. The intelligent system asks a more difficult question: which sequence reduces expected harm?

To answer responsibly, it would need far more than a camera and a list of supplies. It would need verified clinical inputs, time-to-treatment estimates, dependency constraints, inventory states, route conflicts, and clear rules for escalation. It would also need to know when its own information is insufficient. A system that confidently optimizes the wrong objective may be more dangerous than a slower system that defers to a clinician.

This scenario is an illustration of scheduling and triage, not a claim that an AI system should diagnose a child, select a patient for treatment, or overrule medical professionals. Health-related AI

requires documented validation, informed oversight, privacy protection, and an explicit allocation of human responsibility. International guidance emphasizes that human autonomy, safety, transparency, accountability, equity, and sustainability must be built into health AI rather than added after deployment.[9]

The machine's task is therefore not "decide who deserves care." Its task is to assemble an auditable decision space: display the current priorities, estimate the consequences of delay, expose uncertainty, test alternative schedules, and alert authorized clinicians when a threshold is crossed. In a high-stakes system, knowing when not to act is part of intelligence.

The Architecture of an Engineered Mind

A machine that receives data, processes it, and returns an output is not yet an engineered mind. The architecture proposed in this book adds a coordinating layer - metaphorically a brain, but not a biological imitation - that integrates sensory streams, domain knowledge, uncertainty, memory, simulation, and governance. It asks not only "What can be done?" but "What is happening, what matters here, what may follow, what must not be violated, and who remains responsible?"

At the center of this architecture are five living knowledge structures that I call the Five Books. The name is deliberate. Human civilizations have always used books to carry memory across generations. Here, however, a "Book" is not sacred or immutable. It is a versioned, auditable module whose rules, examples, weights, and failures can be inspected and revised.

Book	Engineering role	What it contributes
Watchers	Constraints and safeguards	Defines professional rules, ethical boundaries, permissions, forbidden actions, escalation conditions, and evidence requirements.
Parables	Cases and precedents	Stores worked examples, failures, counterexamples, and the reasons earlier decisions succeeded or failed.
Luminaries	Values and priorities	Represents benefit, harm, fairness, dignity, urgency, and other domain-specific priorities without pretending they are universal constants.
Predictions	Forecasting	Estimates possible future states, their likelihoods, consequences, and sensitivity to missing information.
Dream Visions	Counterfactual simulation	Generates alternatives, rehearses unfamiliar situations, explores creative combinations, and tests what might happen before action.

1. The Book of the Watchers

The Watchers contain the limits that intelligence must not silently rewrite for itself. In medicine, these may include scope-of-practice boundaries, consent, privacy, drug-interaction rules, and mandatory human review. In engineering, they may include safety codes, operating envelopes, redundancy requirements, and shutdown conditions. The Watchers do not make the system wise; they make its authority explicit and contestable.

2. The Book of Parables

The Parables are not moral tales in the literary sense alone. They are structured precedents. Each case records the situation, the available evidence, the chosen action, the expected outcome, the observed outcome, and the lesson. A case can become a positive example, a warning, or a counterexample that prevents the system from repeating an attractive mistake. Unlike a static training corpus, the Parables preserve why a decision was accepted or rejected.

3. The Book of the Luminaries

The Luminaries address the hardest part of any decision: value. The first edition introduced the symbolic relation V = L - H, where value was imagined as the balance between love and hate. That expression captured an intuition, not a validated moral instrument. Love and hate have no universal engineering unit, and a formula cannot absolve a designer of responsibility. In the second edition, the expression remains as a philosophical seed - a reminder that choices often contain attraction and aversion - while operational decisions use defined, measurable terms.

$$\textit{Conceptual seed: } V = L - H$$
$$\textit{Operational score: } S(a) = wB \cdot B(a) - wH \cdot H(a) - wU \cdot U(a)$$

Here, B(a) is an explicitly defined benefit score for action a, H(a) is an explicitly defined harm score, U(a) is uncertainty, and the weights are approved for a specific domain. The result is not “the mathematics of morality.” It is a transparent decision aid. The variables, scales, thresholds, and responsible authority must be documented. If the values cannot be defended, the equation must not be allowed to hide them.

4. The Book of Predictions

Prediction supplies the system with possible futures rather than a single imagined outcome. For every action, the system should estimate what may happen next, what evidence supports that estimate, which assumptions are fragile, and how quickly reality will reveal that the model is wrong. A prediction without a confidence interval or a revision trigger is not foresight; it is merely confidence wearing a scientific costume.

5. The Book of Dream Visions

Dream Visions are the architecture's creative laboratory. They generate counterfactuals: What if a route closes? What if the speaker is afraid rather than angry? What if the familiar proverb is being used ironically? What if the apparent success conceals delayed harm? The system rehearses alternatives without acting on them. Encouragement and penalty can shape exploration, but they should not be described as emotions unless evidence supports that claim. They are control signals used to widen, narrow, or redirect the search space.

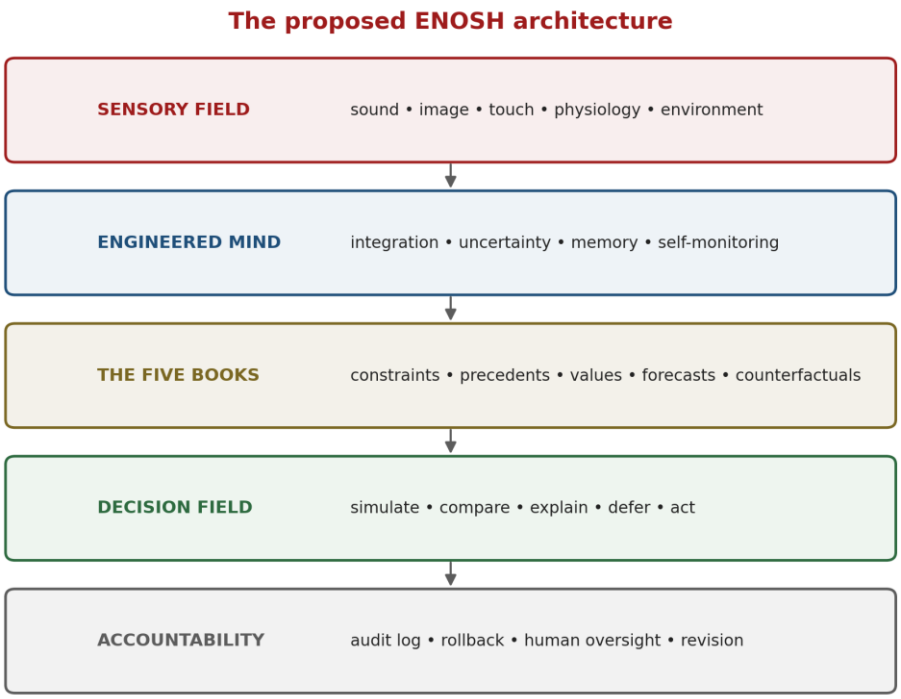


Figure 2. The proposed ENOSH architecture. Sensory inputs are integrated with uncertainty and the Five Books before a decision enters an accountable action pathway.

Resetting Is Not Learning

Earlier experiments encouraged the tempting idea that a dangerous belief could be removed simply by deleting what the system had learned and starting again. A reset may be necessary after corruption, but erasure is not a theory of safety. If the data, objective, or architecture that produced the failure remains unchanged, the same failure can return.

A safer system needs versioned memory, rollback, protected constraints, provenance, audit logs, adversarial testing, and a human-controlled stop mechanism. It should preserve enough history to understand why a correction occurred while preventing disallowed behavior from silently re-entering through new training data. This is governance as engineering. The NIST AI Risk Management Framework expresses the same general principle at an organizational scale: risk must be governed, mapped, measured, and managed throughout the life cycle of an AI system.[10]

From One Artificial Intelligence to a Society of Systems

No serious artificial intelligence will live alone. Medical systems will interact with logistics systems. Educational systems will interact with language, assessment, and accessibility tools. Engineering systems will negotiate with power, climate, security, and maintenance systems. The result will not be one artificial brain but a society of specialized intelligences, each with its own measurements, permissions, and blind spots.

This makes coordination more important than raw intelligence. A medical system should not silently inherit the priorities of a financial optimizer. A scheduling system should not convert urgency into mere throughput. A language model should not treat popularity as truth. Each system needs its own Watchers, Parables, Luminaries, Predictions, and Dream Visions, while a supervisory layer checks conflicts among them. Human authority must remain visible, especially where rights, safety, childhood, health, or irreversible harm are involved.

The "brain" in this architecture is therefore not a single giant model. It is the disciplined integration of specialized processes. It filters, compares, questions, explains, and sometimes refuses. Its most intelligent sentence may be: "I do not have enough evidence to decide, and this is the person who must take responsibility next."

What This Book Claims - and What It Does Not

The central claim of this book is that an artificial system can be engineered to move beyond isolated pattern production toward integrated, uncertainty-aware, value-constrained, and self-correcting decision behavior. Fuzzy logic, multimodal sensing, case memory, counterfactual simulation, and governance can contribute to that movement. Together, they may form a functional precursor to what we call awareness in engineering language.

The book does not claim that fluent output proves consciousness. It does not claim that fuzzy logic creates feelings. It does not treat emotion recognition as emotion, confidence as certainty, a reward signal as desire, or a medical prediction as a clinical diagnosis. Scientific discussions of machine consciousness increasingly stress the need for indicators derived from theories of consciousness, not anthropomorphic impressions alone.[7,8] That discipline protects both science and imagination: we can explore boldly without turning metaphor into evidence.

The examples in this chapter are conceptual and contain no identifiable patient data. ENOSH and any related conversational system are presented here as experimental research architectures, not as autonomous diagnostic instruments. Any future human-participant or clinical evaluation must be governed by appropriate ethics review, informed consent, privacy safeguards, validation, and qualified professional oversight.

The Journey Ahead

The chapters that follow will move from the philosophical question to the engineering structure. We will examine fuzzy and defuzzy reasoning, sensory translation, speech intention, memory, values, simulation, prediction, and the coordination of multiple intelligent systems. The Five Books will return not as decoration, but as working modules whose rules can be defined, tested, criticized, and improved.

The destination is not a machine that imitates humanity perfectly. I do not want an artificial system that merely learns our confidence, our prejudice, and our talent for defending mistakes. I want a system that can preserve uncertainty long enough to examine it; that can hear more than words; that can imagine more than one future; that can explain the principles guiding its action; and that can stop when the responsible human decision has not yet been made.

We began with a cat that was said to have nine lives. The first systems searched for the sentence that most people had repeated. The engineered mind I seek must do something harder: understand why the proverb was used here, what the speaker intended, which interpretation is uncertain, what evidence could change the answer, and whether an answer should be given at all.

The future of artificial intelligence will not be decided by the number of answers a machine can produce. It will be decided by the quality of the questions it has learned not to ignore.

References

1. Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

2. Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies, 7(1), 1-13. https://doi.org/10.1016/S0020-7373(75)80002-2

3. Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, SMC-15(1), 116-132. https://doi.org/10.1109/TSMC.1985.6313399

4. Picard, R. W. (1997). Affective Computing. MIT Press. https://mitpress.mit.edu/9780262661157/affective-computing/

5. Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. Trends in Cognitive Sciences, 10(1), 24-30. https://doi.org/10.1016/j.tics.2005.11.009

6. Beaty, R. E., Kenett, Y. N., Christensen, A. P., et al. (2018). Robust prediction of individual creative ability from brain functional connectivity. Proceedings of the National Academy of Sciences, 115(5), 1087-1092. https://doi.org/10.1073/pnas.1713532115

7. Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? Science, 358(6362), 486-492. https://doi.org/10.1126/science.aan8871

8. Butlin, P., Long, R., Bayne, T., et al. (2026). Identifying indicators of consciousness in AI systems. Trends in Cognitive Sciences, 30(6), 488-501. https://doi.org/10.1016/j.tics.2025.10.011

9. World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. https://www.who.int/publications/i/item/9789240029200

10. Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1