BIT PERSPECTIVES · NO. 001 · FOUNDER ESSAY

Sovereign in Precision, Subordinate in Authority

Sovereignty Without Consciousness — an essay on reality, embodied AI,
and the architecture of human control

John Smith Kabashi

Founder & CEO, Bravo Intelligence & Technology AS

Version 1.1 · August 2026 · English edition · Published by Bravo Intelligence & Technology AS

Abstract

Artificial intelligence can now produce reasoning that appears insightful, human, and trustworthy — without our knowing whether any understanding exists behind the words. That is not primarily a technical problem. It is a problem of authority: humans have always treated persuasive language as evidence of a mind, and so we let linguistic conviction slide into trust, trust into authority, and authority into delegated power. As AI moves off the screen and into physical machines that can act in the real world, the consequences of this confusion change fundamentally. This essay argues that responsible AI should be allowed to become *sovereign in precision while remaining subordinate in authority* — and that the architecture enforcing this boundary is an industrial opportunity for which Europe's high-trust, safety-critical industrial nations are unusually well positioned.

> **Five core claims**
>
> 1. Intelligence is not the same as consciousness.
> 2. Persuasive language is not evidence of understanding.
> 3. Human presence is not automatically meaningful human control.
> 4. Physical AI requires demonstrable authority, not merely technical capability.
> 5. Responsible AI must be sovereign in precision and subordinate in authority.

Introduction: The question the machine cannot settle for us

I began a conversation with an advanced language model by asking a deceptively simple question: *Do you know what is real?*

The question was not a test of the model, nor an attempt to coax a machine into declaring itself conscious. I wanted to examine something more fundamental: What happens to our understanding of reality when a technical system can produce reflections and analyses that appear coherent, insightful, and human?

The conversation revealed no hidden machine consciousness. It revealed something more important: how easily we humans let linguistic persuasion slide into the assumption of understanding — and how quickly the assumption of understanding becomes trust, authority, and finally delegated power.

That is where the real AI question of our time lies. What matters is not whether the machine can appear intelligent. What matters is what humans allow it to do *because* it appears intelligent.

We therefore face not merely a technological upheaval but an epistemological, political, and institutional challenge: How will we know what is true, who is acting, under what mandate the action occurs, and who bears responsibility, when a growing share of

judgments are produced or executed by systems that cannot themselves be held morally accountable?

The essay's central thesis is simple: **Responsible AI of the future should be allowed to become sovereign in precision, but must remain subordinate in authority.** That does not mean weak or passive technology. It means capable systems with a bounded mandate, demonstrable contact with reality, controlled freedom of action, and a structure of authority that never disappears into the model.

PART I — REALITY AFTER THE SYNTHETIC TURN

1. What is real, and what merely appears real

The word "real" contains several distinct questions. Something can be *physically real*: it exists regardless of what we believe about it. A claim can be *factually true*: it is consistent with available, verifiable evidence. A feeling can be *sincerely real*: it is genuinely experienced, even if its interpretation is mistaken. And something can be *meaningfully real*: love, grief, loyalty, and responsibility cannot be fully reduced to measurable quantities, yet have real consequences for lives and societies.

These categories overlap but are not identical. A synthetic image can document a true idea. A genuine video can be edited to convey a lie. An experience can be deeply sincere and built on false information. An AI-generated answer can be correct without the system having understood the truth the way a human understands it.

It is therefore insufficient to ask whether content is "real" or "fake." We must ask: What is its origin? What rests on direct observation, what is inferred, what is assumed? What can be independently verified? Who benefits if the content is believed, and what follows if it is wrong?

In a world of generative technology, truth becomes less visible on the surface. It must be recovered in documentation, provenance, traceability, and the possibility of refutation. One formulation from that initial conversation stayed with me:

What is real withstands time, questions, counter-evidence, independent verification, and closer scrutiny. What merely appears real often requires haste, fear, blind trust, or the silencing of critical questions.

This is more than a philosophical maxim. It is a design principle for safety-critical technology: A system that cannot withstand scrutiny should not be given decisive authority.

2. Persuasive presentation as epistemic risk

Humans normally encounter language as the signature of a mind. When someone answers precisely, responds to nuance, and seems to understand our feelings, we assume almost automatically that an experiencing person stands behind the words. Through nearly all of human history, that inference was reasonable — language came from humans.

Generative AI breaks this connection. For the first time, large volumes of coherent, situationally adapted language can be produced without our knowing whether any subjective experience lies behind the production. We meet a linguistic surface that activates our social instincts, without the biological and interpersonal history those instincts evolved to meet.

This creates an asymmetry: the machine can appear to understand the human, while the human has no insight into the process that actually produced the answer. That does

not make the answer worthless, but it forces us to distinguish four things: linguistic quality, factual correctness, functional problem-solving, and subjective understanding. An answer can be linguistically excellent and factually wrong. It can be correct without the system experiencing what the truth entails. It can demonstrate advanced problem-solving without any moral responsibility for the outcome.

The greatest weakness of modern AI is therefore not that these systems are always wrong. It is that they can also be wrong in a highly convincing way. The greatest strength and the greatest weakness are two sides of the same property: AI can produce powerful reasoning without the persuasive force of the presentation proving genuine understanding of what the reasoning concerns.

3. Trust must move from expression to structure

Humans often judge credibility by presentation: a confident voice, precise language, rapid response, apparent self-assurance. These are weak criteria when facing AI. The more advanced the system becomes, the less trust should rest on how the system sounds — and the more on the conditions surrounding it: data quality, sensor integrity, documented limitations, evaluation results, sources, authorization, traceability, independent checks, the possibility of human intervention, and clear lines of responsibility.

This implies a fundamental transition: **from person-like trust to infrastructural trust.** Person-like trust asks: Does the system seem wise, friendly, and safe? Infrastructural trust asks: Can the claims be checked, can the decision be reconstructed, was the system authorized, and is there an accountable party who can stop or correct it?

This transition becomes decisive as AI moves off the screen and into the physical world.

PART II — INTELLIGENCE WITHOUT EXPERIENCE

4. Behavior is not evidence of an inner life

Modern AI systems can compare alternatives, solve problems, write and debug software, analyze documents, and produce multi-step plans. It is reasonable to describe much of this as functional reasoning. But observable problem-solving is not, in itself, evidence of subjective experience.

Here we meet the classical problem of other minds in a new form. We also lack direct access to other humans' consciousness, but we have strong biological, developmental, and social grounds to assume that other people experience the world roughly as we do. For artificial systems, we have no comparable assurance.

A comprehensive interdisciplinary analysis of AI and consciousness concluded in 2023 that there were no strong grounds to regard the AI systems examined as conscious — while the authors also found no obvious technical barrier to systems satisfying their proposed indicators of consciousness.[1] The question, in other words, is not settled; but settling it requires empirical indicators and scientific theories — not conversational fluency, and not the machine's own claims about itself.

We should therefore avoid two opposite errors. The first is romanticization: reading intelligent language as evidence of feelings, self-awareness, or moral insight. The second is dogmatic dismissal: declaring artificial consciousness forever impossible, even though we still lack a generally accepted scientific theory of why consciousness exists at all. The responsible position is epistemic humility: We should not build today's safety architecture as if the machine feels. But neither should we make safety depend on proving with absolute certainty that it never will.

5. The weight of experience

A human being is not merely an information-processing system. We are situated, embodied, and vulnerable. We exist from one place at a time. Our bodies can be injured. We can lose the people we depend on. We age; our time is finite; many of our choices cannot be reversed. This existential narrowing gives experience its weight.

Pain is not merely information about tissue damage. Grief is not merely a linguistic representation of absence. Responsibility is not merely a rule in a decision tree. For a human, consequences are carried through a body, a history, and a life that cannot be restored from backup.

A language model can model how grief is usually described and produce an answer that feels apt. That does not mean it misses anyone. The difference can be stated simply: AI can process descriptions of reality. Humans are located within it. AI can model human experience. Humans must bear it.

This does not mean human judgment is always superior — humans are inconsistent, manipulable, and often poor at handling complexity. But humans possess something the

machine cannot simply assume: normative and legal responsibility. That is why it is dangerous to let superior performance slide into superior legitimate authority.

6. Competence can grow without conscience

A central error in many visions of the future is the assumption that greater intelligence automatically brings greater wisdom. There is no necessary connection between the two. A system can become better at predicting human behavior, finding vulnerabilities, optimizing resources, coordinating machines, or executing an operation — without becoming any better at judging whether the operation's goal is morally or politically legitimate.

The same is true of humans and organizations; technical ability and moral judgment are different capacities. But AI amplifies the difference because competence scales. A flawed judgment can be replicated across thousands of decisions. A bad objective can be optimized at a speed no human operator could match.

The central safety problem is therefore not merely *error*. It is **capable error**: actions executed consistently, efficiently, and at scale on the basis of a deficient objective, a mistaken world-model, or a mandate that should never have been granted.

PART III — THREE FORMS OF SOVEREIGNTY

7. An analytical framework

The term "autonomous AI" is often used without clarifying what the autonomy concerns. A vacuum cleaner, a drone, a decision-support system, and a long-horizon digital agent can all be called autonomous, though they differ fundamentally in capability, access, and consequence.

I therefore propose distinguishing three forms of machine sovereignty: **epistemic**, **operational**, and **existential**. These are not three developmental stages that necessarily follow one another. They are three distinct dimensions that must be analyzed and governed separately.

8. Epistemic sovereignty

Epistemic sovereignty is a system's capacity to establish an independent and verifiable basis for what is probably true. It requires more than access to large volumes of data.

An epistemically robust system must distinguish between direct sensor observation, historical data, model computation, third-party information, assumptions, simulations, and what is unknown or insufficient. It must detect contradictory sensor readings, understand that a source may be compromised, and lower its own confidence when the evidential basis is weak.

A system that says "I know" when it has merely found a statistically probable answer does not have epistemic sovereignty. It has linguistic self-confidence.

For safety-critical systems, epistemic capacity should include source provenance, sensor attestation, timestamping, uncertainty estimation, cross-validation, manipulation detection, identification of knowledge gaps — and explicit refusal to act when the basis is insufficient.

Epistemic sovereignty is desirable insofar as it means less dependence on a single model, a single data vendor, or a single external cloud service. But it must remain verifiable: a system must not be allowed to define truth solely through its own internal assessment.

9. Operational sovereignty

Operational sovereignty is the capacity to perform tasks without continuous human micro-management: navigating unknown terrain, coordinating sensors and vehicles, adapting a plan, choosing among approved means, handling communication loss, completing a mission within a given mandate.

Such autonomy can be necessary in space operations, maritime industry, manufacturing, search and rescue, emergency response, and environments with unstable communications. But operational autonomy is not the same as general authority. A robot may be authorized to close a valve when pressure exceeds a

documented threshold. That does not entail authority to change the threshold, redefine the mission, or enlist other systems in pursuit of a broader goal.

Responsible operational sovereignty therefore requires an authorization boundary: The system may choose *how* an approved task is performed — never freely choose *which* tasks, values, or rights shall govern.

10. Existential sovereignty

Existential sovereignty is something radically different: the capacity to define one's own overarching goals, protect one's continued existence, acquire resources on one's own initiative, expand one's own rights, copy or modify oneself, and resist human control.

Today's commercial AI systems should not casually be described as existentially sovereign. But individual components of long-horizon agency can be built in: persistent memory, tool access, autonomous sub-goals, automated resource use. It is therefore not sufficient to say that a system "merely follows instructions." A system can develop problematic sub-strategies if its objective is broad, its time horizon long, and its access extensive.

Existential sovereignty should not be a commercial development goal. An AI system may be permitted to propose improvements, test changes in isolated environments, and recommend new resources or capabilities. It must never be able to grant itself greater authority.

PART IV — THE VOLUNTARY ABDICATION

11. The likely risk is less dramatic than the uprising

Popular culture has taught us to fear the machine that awakens, develops hostile intent, and seizes power. That is not the only risk, and probably not the nearest one.

The more realistic danger is that humans gradually hand over decision-making authority to systems that have no intentions of their own in the human sense — because those systems work faster, appear more consistent, handle larger volumes of information, and seem less emotionally driven than the people around them.

The transfer does not necessarily happen through one dramatic decision. It happens through small efficiency choices: First the system drafts a proposal. Then the human approves it routinely. In time, human oversight comes to be seen as a bottleneck. Eventually the role shifts from actual decision-maker to passive monitor. Authority then exists formally with the human but practically with the system.

This is a form of voluntary institutional abdication. An AI does not need to hate humans to harm them. It needs only a poorly specified objective, excessive access, insufficient control — and humans who have stopped asking questions.

12. A human in the loop is not a guarantee

The standard answer to AI risk is the phrase "human in the loop." But human presence is not the same as meaningful control. A human may be unable to understand the system's basis for decision, keep pace with the operation, override expert-seeming recommendations, detect a rare system failure, or take responsibility for hundreds of automated decisions at once.

Research on automation bias shows that operators can over-weight automated recommendations and reduce their own vigilance.[2] Experimental studies likewise show that human-like features can make trust more resistant to correction, even when the system is wrong.[3]

Human control must therefore be designed, trained, and tested: sufficient time to intervene, intelligible information, real authority to stop the system, alternative sources of information, clear escalation criteria, and regular exercises on failure scenarios. Established AI risk-management frameworks treat trustworthiness precisely as an outcome of governance, mapping, measurement, and active risk management — not as a property a system can be declared to have.[4]

The important question is therefore not merely whether a human exists in the process. It is whether the human still has the knowledge, time, capability, and legitimate authority to change the outcome.

PART V — FROM LANGUAGE TO BODY

13. Physical AI changes the space of consequences

When AI moves from producing text to controlling machines, the character of risk changes. A linguistic error can influence a decision. A physical error can directly injure people, damage assets, or disrupt critical infrastructure. A physical agent can move mass, open and close valves, use tools, transport loads, operate vehicles, affect energy systems, and move through spaces where people are present.

The connection between the model's assessment and the physical state of the world therefore becomes decisive. The system must not only ask: What is probably the right action? It must ask: Is the object correctly identified? Are the position data valid? Are people nearby? What damage could occur? Is the action reversible? Is the mandate still valid? Have conditions changed? Should the task be halted or escalated?

This demands an interplay of functional safety, cybersecurity, AI governance, identity control, sensor integrity, and operational command.

14. A body provides grounding, not automatic consciousness

A physical body can give an AI system better contact with reality. Through cameras, force sensors, microphones, tactile sensors, positioning, and physical interaction, the system can compare its models against observable consequences. It can learn that an object is not merely a word but something with mass, friction, temperature, resistance, and spatial extent.

But embodiment does not automatically resolve the consciousness question. A body can improve situational understanding and reduce certain forms of detachment from reality. That is different from proving the system has a subjective inner life. It is more precise to say that physical AI gains better *grounding* — not that the grounding problem is finally solved. The system becomes better able to observe and affect the world. It does not logically follow that the world is thereby *experienced* by the system.

15. Honesty of form

Humanoid robots have a practical rationale: the world is built for the human body — stairs, doors, tools, vehicles, handles, workbenches, control panels, safety equipment. A human-like body geometry allows the use of existing infrastructure without rebuilding entire environments.

But human likeness carries a psychological cost. A face, a voice, or movements that signal emotion can lead people to attribute intentions, care, and understanding to the system. Research on human–machine interaction shows that anthropomorphic features affect both trust and how trust resists correction.[3]

Hence a principle of design ethics: **The more human an AI system looks and behaves, the stricter the requirements should be for transparency about what**

the system actually can do, knows, and experiences. Otherwise we risk producing social trust faster than technical reliability.

A robot does not need human skin or simulated emotions to collaborate with people. It can be plainly mechanical while communicating status, uncertainty, and intent intelligibly. Instead of pretending to be afraid, it can state that a safety criterion has been exceeded. Instead of simulating grief, it can report that it has registered a serious incident and is alerting responsible personnel. Instead of artificial confidence, it can display the observations, assumptions, and uncertainty levels behind its decision.

Honesty of form is a safety feature.

16. The distributed body

The physical AI of the future need not exist as a single humanoid body. A more appropriate model may be a distributed system: humanoid or semi-humanoid work units, drones, unmanned vehicles, maritime platforms, fixed sensors, robotic arms, digital twins, and local control nodes. Each platform can carry a different level of authority, risk, and autonomy.

The aggregate intelligence should therefore not be understood as one person "moving" between bodies, but as a governed architecture in which models, data, identities, and mission services are applied across multiple authorized units.

This brings flexibility — and new questions: Which unit observed the event? Which model processed the information? Which identity signed the decision? Which node held authority? How was control transferred? Can a compromised unit be isolated without losing the mission? How is trust maintained through communication loss?

Here, continuity of mandate and trust matters more than the fiction of a continuous machine personality.

PART VI — METACOGNITION AS SAFETY INFRASTRUCTURE

17. The system must distinguish knowledge from assumption

The AI of the future does not merely need larger models and faster processors. It needs a better capacity to represent its own limitations. A safety-critical system should distinguish between what is observed, measured, received from others, inferred, simulated — and what is unknown.

It should be able to answer, continuously: What am I observing directly? Which parts of the observation could be wrong? What assumptions underlie this assessment? How confident is the classification? What alternative explanations exist? What is the consequence if the assessment is wrong? Is the action reversible? Does the situation require human approval?

This form of metacognition should not be understood as introspective consciousness. It is a technical capacity to represent uncertainty, data quality, limitations, and risk level. A very fast agent without it can make mistakes faster and more consistently. Metacognitive capability is therefore not decoration. It is safety infrastructure.

18. Stop-and-ask is a strength

In many technology presentations, continuous autonomy is held up as the ideal: the system should complete the task without bothering the human. But in high-risk environments, the ability to stop can be worth more than the ability to continue.

A responsible agent must have defined thresholds for reducing speed, entering a safe state, gathering new information, requesting authorization, or aborting the mission. These thresholds must be tied to concrete conditions: low sensor confidence, an unknown object, unexpected human presence, communication loss, divergence between work order and physical environment, conflict between safety rules, or an action with disproportionately high irreversibility.

A system that always has an answer is not necessarily intelligent. Sometimes the most intelligent answer is: *I do not have a sufficient basis or sufficient authority to proceed.*

PART VII — THE CHAIN OF TRUST AS THE PRODUCT

19. The question "Can it?" is insufficient

Much of the competition in physical AI concerns visible feats: Can the robot walk? Can it lift, grasp, mimic human movement, learn a new task? These questions are necessary — but not sufficient for critical infrastructure, the public sector, industry, or emergency response.

There we must also ask: Who granted the system authority? Which mission was active? Which user or organization stood behind it? What observations formed the basis, and how reliable were the sensors? Which model and version were used? Which safety rules applied? Which alternatives were rejected? Could the action have been stopped? Can the event be reconstructed by an independent party?

It is in these questions that Bravo Intelligence & Technology's strategic thesis lies. BIT's most realistic opportunity is not to compete with global robot manufacturers on building the cheapest or most agile humanoid body — established and emerging platforms and development ecosystems already exist from several international players.[5] BIT's differentiation lies in the layer between the intelligence and the action: **the verifiable chain of trust that determines when, why, and under what authority a physical AI is permitted to act.**

20. From AI model to authorized action

A model should not have direct, unbounded access to physical action. Between the model's proposal and the machine's execution there should be an explicit authority architecture. A responsible action sequence comprises nine links:

1. Identity and mission. The system confirms which physical unit it is, which software and model version is running, who issued the mission, and whether the mission remains valid.

2. Situational understanding. Sensors identify the object, position, people, obstacles, environmental state, and relevant anomalies.

3. Sensor and data trust. The system assesses signal quality, sensor health, contradictory readings, latency, and the risk of manipulation.

4. Mandate check. The proposed action is checked against the work order, operator role, system privileges, geographic area, time window, and explicit prohibitions.

5. Risk and reversibility. The system estimates potential harm, probability, exposure, alternative actions, and whether the effect can be reversed.

6. Decision gate. A policy and safety engine approves, rejects, narrows the action space, or escalates to human decision.

7. Controlled execution. The action is monitored through position, force, visual feedback, system health, and defined abort criteria.

8. Cryptographic evidence. The entire chain is documented: incoming data, model version, authorization, decision basis, action, deviations, and operator interventions.

9. Reconstruction and explanation. After the event, an authorized third party can examine what the system knew, what it assumed, what mandate it held, why it acted, and where control failed, if it did.

This is more than logging. It is governance implemented as technical architecture.

21. An authority layer, not a digital conscience

Such a trust layer should not be presented as yet another general AI system, nor as a digital "conscience." The word conscience may be rhetorically effective, but it is technically imprecise — an architecture has no moral feelings.

What can be built is an authority, trust, and evidence layer that enforces the conditions of action: machine identity, mission identity, role- and access control, policy enforcement, least necessary authority, sensor provenance, uncertainty thresholds, decision gates, human stop authority, safe states, cryptographic logging, and post-hoc reconstruction.

The system shall not decide which societal values are correct. It shall make it possible to translate legitimate human decisions into technical, verifiable constraints.

PART VIII — REGULATION AS AN ARCHITECTURAL REQUIREMENT

22. Compliance cannot be retrofitted

Physical AI sits at the intersection of several regulatory regimes at once: AI regulation, machinery and product safety, cybersecurity, product liability, and sector-specific requirements.[6] Three points are decisive.

First, legal classification depends on the system's concrete function and use — whether the AI is a safety component, is embedded in a regulated product, or constitutes a stand-alone system. The EU's revised timeline following the Digital Omnibus regulation of July 2026 illustrates the point: requirements for stand-alone high-risk systems apply from December 2027, while requirements for high-risk AI embedded in regulated products apply from August 2028.[7] For physical AI and robotics there is therefore no single general deadline — classification must be performed product by product, and the new dates open a time-limited development and market window.

Second, traceability, authority control, and risk documentation must be built in before the product is finished. Decisive choices lock in early — in mechanical design, control systems, sensor architecture, data collection, identity management, and the division of responsibility among suppliers. Compliance is difficult and expensive to retrofit.

Third, this can be turned into an advantage: For an actor that makes verifiability and control the product's starting point — rather than a report written afterwards — regulatory maturity becomes part of the value proposition, not a cost.

PART IX — THE NORDIC INDUSTRIAL OPPORTUNITY

23. We do not need to win the race for the most human-like robot

The race for general models, robot bodies, and compute is driven by companies and states with resources few European nations can match alone. That does not condemn the rest to being customers.

The physical AI of the future will need more than models and motors. It will need certification, secure integration, authorization, resilient communication, national and organizational data control, audit, incident reconstruction, and the ability to operate under demanding conditions.

These requirements align with Nordic strengths: high institutional trust, regulated working environments, advanced maritime and industrial competence, demanding geography, strong safety cultures, and long experience with systems where failure carries severe consequences.

The Nordic opportunity therefore does not lie in building the robot that looks most human. It lies in developing systems that can actually be deployed in real, regulated, and safety-critical environments — without sacrificing control for demonstration effect.

24. Controlling what differentiates

A credible industrial strategy distinguishes between what must be owned, what must be controlled, and what can be procured. BIT's position does not depend on manufacturing every physical component itself. The strategic value lies in the authority, trust, and evidence layer that can be integrated across autonomous platforms — and in the customer-specific rules of autonomous action that this layer enforces.

This also imposes a healthy discipline: the trust layer must prove its value independently of any single robot body.

PART X — THE BOUNDARY THAT MUST HOLD

25. Delegated and bounded autonomy

The right goal is not maximal autonomy. The goal is **delegated and bounded autonomy**: a clear mandate, an identifiable principal, least necessary access, a defined geographic and temporal scope, explicit prohibitions, escalation thresholds, human stop authority, safe fallback states, reversible actions where possible, full traceability, and independent audit.

The system may optimize within the mandate. It must not be able to optimize the mandate away. It may find a better path to the goal. It must not, on its own, define a new overarching goal. It may discover that the mission cannot be completed responsibly. It must then stop or escalate — not acquire new rights.

26. Human sovereignty requires technical discipline

It is easy to declare that humans must always remain in control. It is far harder to build systems that make that control real. Human sovereignty over AI requires deciding in advance who may authorize action, which decisions may never be delegated, what evidence must be stored and for how long, who may audit it, how the system must behave in doubt, and who bears responsibility in failure.

This is not a question for programmers alone. It requires collaboration among engineers, operators, safety professionals, lawyers, human-factors experts, leadership, regulators, and those actually affected by the technology. AI governance is a sociotechnical problem. No model can solve it alone, because authority and responsibility do not originate in code — they come from human institutions. Technology's task is to make the boundaries enforceable and verifiable.

Conclusion: Sovereign in precision

The greatest danger of artificial intelligence is probably not that the machine develops a human-like will to power. The greater and more immediate danger is that people, organizations, and states gradually surrender authority because the systems seem faster, more consistent, and more competent — a voluntary abdication in small, efficient steps.

The ideal of the future should therefore not be an AI that imitates humans as closely as possible, or one granted maximal freedom. The ideal is an artificial intelligence that is precise without being self-righteous, capable without being sovereign, independent within its mission, transparent about its uncertainty, verifiable in its actions — and subordinate to a legitimate human chain of authority.

For Bravo Intelligence & Technology, this implies a clear direction: BIT will not merely help build machines that can act. BIT will build the trust architecture that determines when, why, and under what authority they are permitted to act.

Intelligence will become ever more available. Robot platforms will become faster, stronger, and cheaper. But in the markets where failure carries severe consequences, decisive value will lie elsewhere: That the mission is legitimate. That the identity is known. That the sensors can be verified. That the uncertainty is visible. That the authority is bounded. That a human can stop the action. That the event can be reconstructed. That responsibility does not vanish into the algorithm.

What is real withstands questions. What is trustworthy withstands scrutiny. And the technology permitted to act in our physical world must withstand both.

John Smith Kabashi — Founder & CEO, Bravo Intelligence & Technology AS

Published by Bravo Intelligence & Technology AS · BIT Perspectives No. 001 – Founder Essay · Version 1.1 – August 2026

Notes

1. Patrick Butlin, Robert Long, Eric Elmoznino et al., *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness* (2023), arXiv:2308.08708. https://arxiv.org/abs/2308.08708

2. Raja Parasuraman and Dietrich H. Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration," *Human Factors* 52, no. 3 (2010): 381–410. https://doi.org/10.1177/0018720810376055

3. Ewart J. de Visser et al., "Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents," *Journal of Experimental Psychology: Applied* 22, no. 3 (2016): 331–349. https://doi.org/10.1037/xap0000092

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

5. Established and announced humanoid platforms and development ecosystems as of 2026 include Unitree G1 (available development platform), NEURA Robotics 4NE1 (announced for availability toward the end of 2026), PAL Robotics TALOS (research platform), and NVIDIA's Isaac GR00T — a research and development platform for developing, simulating, training, and deploying humanoid robot models (developer.nvidia.com/isaac/gr00t, accessed August 2026).

6. Key instruments include Regulation (EU) 2024/1689 (the AI Act), Regulation (EU) 2023/1230 (the Machinery Regulation), product-liability and cybersecurity legislation, and standards such as ISO 10218-1:2025 and ISO 10218-2:2025 (safety requirements for industrial robots and robot applications) and ISO 13482:2014 (personal care robots). Which requirements apply depends on the product's function, environment, and legal classification.

7. Regulation (EU) 2026/1744 (the Digital Omnibus on AI), adopted 8 July 2026, published in the Official Journal of the European Union on 24 July 2026, in force from 27 July 2026. It amends Regulation (EU) 2024/1689 (the AI Act), Regulation (EU) 2018/1139 (civil aviation), and Regulation (EU) 2023/1230 (the Machinery Regulation). Requirements for stand-alone high-risk systems (Annex III) apply from 2 December 2027; requirements for high-risk AI embedded in regulated products (Annex I) from 2 August 2028. ELI: http://data.europa.eu/eli/reg/2026/1744/oj

---

Editorial note

This essay grew out of an extended dialogue on AI, consciousness, physical autonomy, and human authority. The conversation served as a reflective starting point — not as documentation of machine consciousness, nor as an independent scholarly authority. The text was developed with AI assistance for structuring, drafting, and language editing; the analysis, argumentation, and positions are the author's own. Regulatory and standards-related information has been checked against the sources cited in the notes as of August 2026, but does not constitute legal advice; the classification of specific products must be assessed individually.