THE JBBA

ESSAY

OPEN ACCESS
ISSN Online: 2516-3957
ISSN Print: 2516-3949
https://doi.org/10.31585/jbba-9-2-(3)2026

Cognitive Sovereignty and Pre-observational Judgement: A Research Agenda and Policy Framework to Safeguard Original Thought in the Age of Machines

Naseem Naqvi MBE
Centre for Evidence-Based Blockchain, The British Blockchain Association, UK

Correspondence: cebb@britishblockchainassociation.org
Received: 20 January 2026 **Revised:** 14 April 2026 **Accepted**: 17 April 2026 **Published**: 19 April 2026

Abstract

Artificial intelligence is quietly reshaping human cognition. As the cost of generic knowledge work falls to near zero, the premium on genuinely original thought rises sharply. The defining risk of the present moment is not that machines will out-think us, but that humans will gladly surrender thinking itself. Drawing on the evidence-based tradition and a decade of institutional work in blockchain governance, this essay introduces the concept of 'pre-observational judgement': the capacity to decide what is worth noticing before any data exists. This faculty is the oldest form of human intelligence and the most resistant to automation, because it operates before the data on which a machine could be trained. From this follows a hypothesis: artificial intelligence, as currently conceived, will struggle to produce original ideas of civilisational importance, because a system trained on what is known cannot be the source of what is not yet thinkable.

The paper then develops a three-part argument on the stakes of this moment: ethics that cannot be reduced to rules, the authorship of the questions that govern collective life, and the preservation of intellectual priority in an era of synthetic abundance. It repositions blockchain as an epistemological infrastructure. Artificial intelligence collapses authorship into probability; blockchain restores it as a matter of record.

The essay closes with six programmes of research and institutional reform: an empirical test of the central claim; a sustained study of pre-observational judgement across industrial, scientific, and legal domains; a blockchain-anchored pilot for the verifiable attribution of original thought; a longitudinal comparison of doctoral training under conditions of AI augmentation; a National Philosophy and Cognitive Sovereignty Framework; and a governance model for embedded philosophers in AI research. These are offered as testable commitments, open to refinement and refutation.

Keywords:
Artificial Intelligence, Blockchain, Philosophy, Epistemology, Pre-observational Judgement, Evidence-Based Blockchain, Decentralisation

JEL Classifications: *A12, B41, D83, O33, O34*

The Philosopher's Hour

There is a quiet shift happening in how human beings think, and it has arrived faster than any curriculum or policy framework has been able to anticipate. The greatest danger of artificial intelligence is not that it will think for us. It is that we will forget that thinking was ever our responsibility. This begs the question: Will the future belong to those who use machines best or will it belong to those who remain capable of thinking without them?

I have been watching this happen in real time. Colleagues delegate their reading to summary engines. Students generate dissertations in an afternoon. Executives produce strategy papers by typing three sentences into a prompt box. A generation of bright young minds is being quietly trained to believe that the point of thought is to arrive at an output, and that any path to the output will do. The temptation is not that machines will replace thinking. It is that humans will gladly surrender it. This is a civilisational mistake. And correcting it will, I suspect, require the most unlikely of heroes: the philosopher.

The paradox of abundance

We have never had more information and never used less of it well. The library of every great civilisation now fits in a pocket, yet original ideas are becoming rarer, not more common. The reason is not that human beings have grown stupid. It is that we have confused two different activities.

The first is the retrieval of what is already known. Machines are excellent at this, and rightly so. The second is the generation of what is not yet known: the formation of new hypotheses, the noticing of what nobody else has noticed, the framing of questions that change how later people see the world. This second activity has never been the property of libraries or search engines or language models. It has always been the work of the observant individual, sitting with a problem long enough to see through it.
We are generating answers at a scale never seen before, while quietly losing the capacity to ask original questions. This is not a technological failure. It is a philosophical one.

I argued in a recent essay [1] that ancient philosophers understand blockchain better than most of us do. The reason is that they thought for long enough about trust, truth, memory, and consensus, and that their work quietly predicted the intellectual conditions under which such technologies would one day become necessary. They sat with questions while we skim answers. Aristotle's insistence that philosophy begins in wonder, Plato's warning in the Phaedrus that the technologies of recording could hollow out the practice of thinking, the Stoic discipline of prosoche or disciplined attention, the Hippocratic tradition of evidence-based observation, and the Socratic willingness to be publicly wrong. Each addresses, in embryonic form, a problem this essay encounters in modern dress.

What machines cannot do

There is a fashionable view that the modern language model is a kind of synthetic mind, and that we are on the cusp of machines that will out-think us. I work in two different fields (healthcare and blockchain) where I see both the promise and the limits of technologies every week. Let me state the matter plainly. This concept extends a tradition that runs from Kant's conditions of possibility through Polanyi's tacit knowledge and Hanson's theory-ladenness of observation, but specifies something those accounts do not isolate: the faculty of deciding what is worth attending to before any observation begins.

The defining difference between a mind and a machine is not intelligence. It is the capacity to decide what is worth noticing before any data exists. Call this faculty **pre-observational judgement**. It is the oldest form of human intelligence and the one most resistant to automation, because by definition it happens before the data on which a machine could be trained. This is the point at which a civilisation decides what it is willing to notice, and therefore what it is capable of discovering.

Machines do not critically observe. They receive input. Critical observation is value-laden attention: the disciplined act of deciding, before any measurement begins, what is worth noticing in the first place. In medicine I learned early that the best clinicians are not those who have memorised the most facts. They are those who notice the subtle signs, who form an original hypothesis and test it against reality. I saw a patient recently whose symptoms did not match any textbook presentation I could recall. Sitting with that mismatch, letting it trouble me across the week, is what eventually produced the right diagnosis. A machine, fed the same notes, would have confidently matched the nearest pattern and moved on. It cannot be troubled by a fact. It cannot be surprised.

Machines do not form hypotheses. They interpolate. The difference between a well-trained model and an original thinker is the difference between a mirror and a window. A mirror shows you what already exists with high fidelity. A window shows you something new. The great breakthroughs of modern science, from penicillin to general relativity, from the DNA double helix to the invention of Bitcoin, and most recently to CRISPR, all began as anomalies rather than aggregations. Each was a detail a machine averaging the literature would have flagged as noise and discarded. Pattern-matching does not notice outliers, by definition. A mind that has decided to attend to what does not yet fit does. The innovations are not always hidden in the patterns; sometimes they are hidden in the exceptions.

Artificial intelligence, as currently conceived, will struggle to produce novel ideas of civilisational importance. It will accelerate our discovery of what we already half-suspect. It will compress centuries of existing thought into usable form. It will not produce the thought that changes what a civilisation suspects in the first place. Because originality does not emerge from pattern recognition. It emerges from the decision to look where no pattern yet exists. A system trained on what is known cannot be the source of what is not yet thinkable.

Machines do not wonder. Wonder is the engine of every great tradition of thought, and it cannot be compressed into a weight matrix. A machine does not stand beneath the stars at night and feel the old shiver of not-understanding that produced astronomy. It does not walk through a hospital and feel the moral weight of what it has seen. Without wonder there is no question. Without the question there is nothing for even the cleverest instrument to answer.

Evidence-based practice, which I have championed all my professional life [3], is not a bureaucratic procedure. It is a philosophical posture. It begins with the disciplined eye of the observer, proceeds through original hypothesis and rigorous experiment, and ends in honest revision. Machines can test hypotheses at a speed we could not match. They cannot originate them. They cannot identify which pattern in the universe is worth pursuing in the first place.

The deeper threat is not that machines will be wrong. It is that humans will stop being willing to be wrong. Willingness to be wrong, publicly, on the record, under the scrutiny of one's peers, is the beating heart of every serious intellectual tradition. Without it, science becomes theatre, scholarship becomes performance, and the whole apparatus of human knowledge quietly ceases to self-correct. That courage is ours alone to preserve. A system that cannot admit error cannot discover truth. It can only refine illusion.

The coming scarcity

Here is a forecast I am prepared to defend: Within a decade, the most sought-after human beings in any organisation of

consequence will not be those who can operate the machines. It will be those who can do what the machines cannot.

They will be the people who ask better questions. They will be the people who can tell when a fluent, confident answer is wrong. They will be the people whose training has taught them to sit with difficulty, to distrust easy consensus, to notice the anomaly that everyone else is skimming past. In short, they will be philosophers in the old and serious sense of that word. Not people with a particular academic credential, although many will have one, but people whose habits of mind are genuinely their own.

Leading artificial intelligence laboratories have already begun hiring philosophers [2] to help them think about machine consciousness, and major research centres are quietly recruiting scholars of ethics, epistemology, and the philosophy of mind. When a technology becomes capable enough that its unintended consequences pose civilisational risk, the people you need are not more engineers. You need people who can think about what the technology is, what it should be, and what it must never be. That is philosophy. It is not a luxury. It is the discipline that keeps the rest of the disciplines honest. Philosophers, properly understood, are the custodians of thinking itself.

This is why a national AI policy without a national philosophy policy is half a strategy. Countries spending billions on computational capability without also spending on the slower, less visible work of producing a generation of citizens who can still think without the machines they are building will discover, a decade from now, that they have bought the hardware and lost the mind. The future is not decided by those who have the most answers. It is decided by those who control the questions. A civilisation that stops asking its own questions does not merely lose curiosity. It loses sovereignty. And once sovereignty is lost at the level of thought, it cannot be recovered by politics. AI shapes the answers. Philosophers must shape the questions.

Why the PhD will matter again

For years, the PhD has been under quiet ridicule. Too long, too expensive, too narrow. An indulgence in a world that rewards speed. I predict this will reverse, and quickly.

The PhD, at its best, is not a qualification. It is a controlled environment for intellectual failure. It trains a mind to formulate a question nobody has properly asked, to design an experiment that could genuinely disprove what one hopes is true, to sit with negative results and let them change one's mind, and to submit one's work to people whose job is to destroy it. Nothing in the training of a machine approaches this. A language model cannot produce a falsifiable hypothesis, defend it against a hostile committee, revise it under pressure, and publish it under its own name. Only a thinking being can do that, because only a thinking being has something to lose.

As the cost of generic knowledge work falls to near zero, the premium on genuinely original knowledge will rise sharply. The doctoral tradition is not obsolete. It is the pedigree from which the next generation of serious minds will emerge. It cannot be mass-produced, and it cannot be synthesised. It has to be grown, one thinker at a time, by other thinkers.

I suspect the next generation of universities will quietly divide into two kinds. The first will deliver AI-assisted education at scale, producing fluent operators of systems they do not fully understand. The second will retain the older work: the slow, adversarial, human-to-human cultivation of original minds. Societies that fail to maintain the second will discover, a generation from now, that they have no one left who can think for them.

The renaissance of original voices: Why original writers and thinkers will be indispensable

Anyone can produce text now. This is the very reason that real writing will become more valuable, not less. When the river is flooded, clean water becomes precious.

The writer who still writes from first-hand experience, who has read the primary literature rather than the summary of it, who has formed a position and is prepared to defend it in public, and who has a story to tell, will stand out with almost embarrassing clarity against the background hum of synthetic prose. The same is true of the speaker. A genuine speaker does not merely transmit information. They transmit presence. They carry the trace of a mind that has wrestled with reality and come away changed. Audiences already sense this, even when they cannot articulate it. As machine-generated content becomes ubiquitous, they will hunger for the voice that carries the scent of lived truth.

There is a deeper point here. A civilisation depends on the existence of people who are willing to say, publicly, "I have thought about this, I have studied this, and here is my view." Without such people, a society loses the capacity to correct itself. It drifts on the current of whatever opinion is most easily produced at scale. The original writer and the original speaker are not ornaments of culture. They are its immune system.

Ethics is not a decoration

The argument that follows turns on three things that machines cannot do for us and must not be allowed to do in our place. The first is ethics. The second is the authorship of the questions that govern our collective life. The third is the preservation of who first thought what. Each of these requires a human being, and each becomes more urgent, not less, as the machines become more capable.

Every serious ethical tradition I know of agrees on at least this much. Morality is not a set of rules bolted onto action after the fact. It is the shape of the action itself, worked out from a careful understanding of what human beings are, what they owe one another, and what it is for a life to go well or badly.

This is why outsourcing our ethical thinking to machines is a category error. A machine can be given rules. It cannot be given a conscience, because a conscience is not a rule. It is the living capacity of a person to feel the moral weight of what they are doing, in context, with all its particulars, and to act accordingly. No amount of reinforcement learning produces this. You can

train a system to avoid a proscribed output. You cannot train it to care.

As we hand more of our decisions to systems that do not care, we are going to discover, painfully, that care cannot be optional. What is the right trade-off between efficiency and dignity when dignity has no line of code? When does a personalised service become a manipulation? What does fairness look like in a probabilistic model that has no lived experience of suffering? Who is accountable for decisions made by a machine? These are not engineering problems. They are philosophical ones embedded within technological systems. Every generation has had to answer versions of them. Ours will have to answer them with stakes higher than any before.

There is a harder-edged version of this argument that needs to be said out loud. If we surrender the authorship of our questions to machines as well as the authorship of the answers, we will find ourselves governed, quietly and effectively, by outputs whose origins we can no longer trace and whose assumptions we can no longer audit. This is not a hypothetical risk. It is the slow working out of what it means to delegate thought. Without a serious ethical framework, intelligence becomes dangerous. The people best equipped to help us build such a framework are those who have spent their lives thinking about exactly this, and they are, once again, philosophers.

Blockchain and the preservation of original thought

I have spent the better part of a decade building the institutional infrastructure of blockchain (in this country, and globally) and I want to say something about why it matters for the argument I am making.

Original thought is fragile. It can be plagiarised, rewritten without attribution, ingested into training sets without consent, or quietly erased from the record. A civilisation that cannot reliably say who first said what, and when, loses something far more important than intellectual property. It loses its memory. Without that memory, the incentive to produce original thought collapses, because the reward structure collapses with it.

AI produces plausible answers without origin. Blockchain records origin without distortion. One optimises for usefulness. The other preserves truth. Blockchain is the first technological system in history that treats truth as something which must be historically anchored rather than merely statistically inferred. Artificial intelligence collapses authorship into probability. Blockchain restores authorship as a matter of record. It insists that the individual voice, timestamped and verifiable, still matters.

This is why, in an era of synthetic abundance, a credible, tamper-resistant record of intellectual priority is one of the few things left that can protect an original thinker. If a philosopher writes a genuinely new idea in 2026, and that idea is ingested, reworded, and redistributed at scale by systems that do not credit her, she has lost something real. Not only income. Something closer to the moral point of having thought the thing at all.

Evidence-based practice gains from this as well. Hypotheses, experiments, and observations can be recorded immutably, creating a transparent chain of reasoning that is resistant to retrospective editing. This is more than technological innovation. It is a philosophical infrastructure for truth. The technologies of verifiable record and the traditions of original thought belong together. The first preserves the second. Without preservation, originality starves.

The pedigree that cannot be synthesised

There is a quiet fact about genuine thinkers that is rarely discussed. They are often produced by other genuine thinkers. You have to be in the room (or with their work) over years, with someone whose habits of mind are themselves original. You watch how they read and how they react to a weak argument. You watch how they change their mind when the evidence demands it. You absorb, slowly, the temperament of serious thought.

This is a pedigree in the old sense of the word. It is not a genetic matter. It is a transmission of intellectual character from one generation to the next, and it happens person by person. Philosophers create philosophers. And when a generation fails to do so, the loss is not immediate. It is discovered too late. Machines cannot replace this, because the thing being transmitted is not information but a way of being with a problem.

Every civilisation that has produced great thought has understood this. The ancient schools, the medieval universities, the early modern academies, all were built around the idea that a serious mind is grown rather than manufactured. If we allow the machines to hollow out the institutions in which this transmission happens, we will not be able to rebuild them on demand. It takes generations to produce a generation of thinkers. We should protect the conditions under which it happens, with some urgency.

The architecture of education, and the recovery of wonder

If we train the next generation only to prompt machines effectively, we will produce skilled operators, not free minds. Education will have to recover something it has been quietly neglecting for a generation. The place of slow thinking. The value of observation. The dignity of sitting with a question without an answer. The habit of intellectual humility in the face of machine confidence, which rarely expresses doubt. And above all, wonder. In a world of instant answers, the capacity to stand in awe before the mystery of existence becomes, oddly, a revolutionary act. It is from that disposition that the best questions arise, and without better questions no amount of computation will save us. A student who never struggles with a question will never own an answer.

Why the machines will come to the philosophers

I want to close with a claim that may sound odd, but which I believe to be true. Eventually, the machines themselves will need us. Not for the routine tasks of life, which they will do better than we do, but for the questions that the routine of their operation will force into view. What problems should be

prioritised? What values should be embedded? What trade-offs are acceptable, and at what point does a trade-off become a betrayal? These are not engineering decisions. They are philosophical ones, and only people formed by a serious tradition of thought will be able to answer them.

The builders of these systems are already discovering this. The deep irony of the age we are entering is that as machines become more capable, the need for human wisdom becomes more acute, not less. The question is whether our society is producing the philosophers they will need to consult, in sufficient number and quality, to meet the moment. My worry is that it is not. My hope is that it still can. Within a generation, the most valuable individuals will be those whose thinking cannot be simulated.

Why this debate matters now more than ever before

We stand at a crossroads. One path leads to a world of synthetic consensus, where agreement is generated before understanding, where decisions are optimised but not understood, where knowledge is abundant but wisdom is scarce. The other leads to a renaissance of thought, where technology amplifies human capability without replacing human inquiry, where originality is valued, where ethics guides innovation, and where evidence-based practice anchors progress. The choice is not technological, but rather philosophical.

The stakes are threefold. Ethics, because a machine can be given rules but not a conscience, and care cannot be optional in a civilisation that still wishes to call itself humane. Sovereignty, because the future is decided by those who control the questions, and a civilisation that stops asking its own questions loses more than curiosity. And attribution, because original thought is fragile, and an economy that cannot reliably say who first said what will watch its best minds either give up or leave. This is where blockchain, too often treated as a financial technology, becomes an epistemological one.

We need more philosophers trained in the old sense. More evidence-based practitioners willing to observe, hypothesise, and be corrected. More writers and speakers willing to say what they think and defend it in public. We need institutions, including those built on new technologies such as blockchain, that protect the record of who first thought what. We need an ethical culture serious enough to govern the machines before the machines quietly govern us. The future will not belong to those who use machines best. It will belong to those who remain capable of thinking without them.

Future work and recommendations

The arguments set out in this essay offer a diagnosis. The work that follows must be done. I propose here six programmes of research and institutional reform that would test, refine, and extend the claims made above. Each is offered not as a polished design but as an opening statement, published in this Journal so that others may take it up, improve it, or refute it.

1. Empirical testing of the central claim

The abstract of this essay proposes a hypothesis: that artificial intelligence, as currently conceived, will struggle to produce original ideas of civilisational importance. That hypothesis deserves to be tested. I propose a prospective, decade-long study in which a blinded corpus of research proposals, drawn from senior human researchers and from frontier AI systems, is assessed by panels of scholars across disciplines against a pre-registered definition of originality. The British Blockchain Association, in partnership with academies of science and the humanities, is well placed to host the protocol. The result, whichever way it lands, will be more useful than another decade of opinion.

2. Operationalising pre-observational judgement

The central concept of this essay warrants a sustained research programme of its own. How is pre-observational judgement acquired? Can it be taught, and if so, how? Is it degraded by heavy reliance on machine-generated outputs? I propose cross-disciplinary studies in clinical medicine, experimental science, and legal reasoning to identify its behavioural signatures and to measure its preservation under varying conditions of technological augmentation. The aim is a taxonomy of judgement that can inform curriculum design, professional training, and the design of AI-assisted work environments.

3. Blockchain-anchored attribution for original thought

The essay argues that blockchain, properly understood, is an epistemological infrastructure for the preservation of original thought. I propose a pilot programme, hosted by the British Blockchain Association in partnership with selected academic publishers, to timestamp and cryptographically attribute scholarly contributions at the point of ideation, not merely the point of publication. Over a five-year horizon, the pilot would measure adoption rates, citation resilience, and resistance to uncredited ingestion by AI systems. If successful, it becomes the template for a standards framework on intellectual priority in the age of synthetic abundance.

4. Doctoral training in the age of machines

The essay predicts that universities will divide into institutions that deliver AI-assisted education at scale and those that retain the slow, adversarial, human-to-human cultivation of original minds. This prediction is testable. I propose a longitudinal comparative study of matched doctoral cohorts across the two models, measured on originality of contribution at five and ten years post-viva, with originality assessed independently by expert panels and by long-term citation patterns. The result will either vindicate the doctoral tradition or require its defenders to update their arguments.

5. A National Philosophy & Cognitive Sovereignty Framework

The essay argues that a national AI policy without a national philosophy policy is half a strategy. I propose the development of a concrete instrument: a National Philosophy and Cognitive Sovereignty Framework, to sit alongside existing national AI strategies. Its elements would include ring-fenced doctoral funding in philosophy of mind, epistemology, and ethics; a national fellowship scheme for embedded philosophers within publicly funded research institutions; and mandatory philosophical review of major public investments in artificial

intelligence. I invite policy makers, academics, and think-tank bodies to author such a framework.

6. Ethics embedded, not appended
The recent appointment of in-house philosophers at leading AI laboratories signals a shift from ethics as compliance to ethics as design. I propose that this practice be studied, codified, and spread. Specifically, I call for a cross-industry working group to develop a governance model for embedded philosophers in AI research, addressing scope of authority, publication rights, independence from commercial pressure, and integration into pre-observational design decisions. Done well, this is the philosophical equivalent of the chief scientific adviser. Done poorly, it is window-dressing. The distinction will be made by governance, not by sentiment.

These recommendations are offered in the spirit of evidence-based practice: as testable commitments, open to refinement and refutation. They stake no authority; they issue an invitation to scholars, policymakers, and practitioners, to carry this programme further.

The future will not belong to those who use machines best. It will belong to those who remain capable of thinking without them, and to the institutions with the courage to protect them.

Competing interests:
None declared.

Ethical approval:
Not applicable.

Author's contribution:
NN drafted this essay in its entirety. The arguments, structure, and conclusions of this essay are the author's own.

Funding:
None declared.

Acknowledgements:
The author thanks colleagues at the British Blockchain Association, particularly at the Centre for Evidence Based Blockchain, for their critical reading of earlier drafts; members of the UK All-Party Parliamentary Group on Blockchain Technologies for discussions that shaped the policy debate; and the anonymous reviewers of this Journal whose challenges sharpened the argument. The essay's deepest intellectual debts, however, are to all the philosophers who have ever lived: to Aristotle on wonder as the origin of philosophy, to Plato's Phaedrus on the fragility of memory, to the Stoic discipline of prosoche, to the Hippocratic commitment to evidence-based observation, and to the Socratic willingness to be publicly wrong. Each of these minds helped shape the narrative of this essay.

References

[1] N. Naqvi, "Why Ancient Philosophers Understand Blockchain Better Than Most of Us Do," The Journal of the British Blockchain Association, vol. 9, no. 1, 2026, doi: https://doi.org/10.31585/jbba-9-2-(1)2026

[2] Google DeepMind Hires a Philosopher to Prepare for Machine Consciousness: https://x.com/dioscuri/status/2043661976534950323

[3] N. Naqvi, "Evidence-Based Blockchain: Findings from a Global Study of Blockchain Projects and Start-up Companies," The Journal of The British Blockchain Association, vol. 3, no. 2, pp. 1–13, Aug. 2020, doi: https://doi.org/10.31585/jbba-3-2-(8)2020.