

# The Battle for Common Sense in the 21st Century: Is Humanity's Last Bastion Falling?

Ramin Saadat

---

AI Disclosure and Translation Statement:

The original manuscript of this work was authored in Farsi (Persian) by the researcher. The English version was produced through AI-assisted translation and linguistic refinement, utilizing Google Translate and Large Language Models. The author has meticulously reviewed, edited, and verified the final English text to ensure that all theoretical nuances, philosophical arguments, and conceptual frameworks accurately reflect the original Persian content. The author remains fully responsible for the integrity and accuracy of the translation. The original Persian manuscript is available from the author upon reasonable request.

---

## Abstract

The concept of "common sense", understood as the body of self-evident and shared beliefs by which people navigate everyday life without the need for explicit reasoning, has entered a new crisis in the twenty-first century. This article argues that, contrary to the philosophical tradition, the question today is no longer what common sense is or how long it will survive. The question is who constructs it, at what scale it is distributed, and whether its nature remains human.

The article's theoretical framework holds that common sense has never been an intrinsic property of the individual mind, but always a social technology: a tool that societies have developed to reduce cognitive costs and enable collective coordination. Like any technology, it can be designed, dismantled, or rebuilt.

The article first examines four intellectual currents of the past quarter-century: the analytic epistemological defence of basic beliefs; attempts to model and artificially reproduce common sense; the diagnosis of the collapse of the shared social infrastructure of reality; and nascent efforts to reconstruct that infrastructure through information ethics and digital governance.

It then shows that these currents are engaged in three fields of struggle: the battle over ownership (the transfer of common sense from the human mind to corporations, states, and machines); the battle over territory (the tension between global algorithmic homogenisation and the need for planetary thinking); and the battle over identity (whether common sense remains human or becomes algorithmic).

The article's central argument is that, faced with "thinking machines" capable of justifying any claim through apparently logical reasoning, human common sense remains the last line of defence against false and unwarranted beliefs. This is due to its rootedness in shared experience and its capacity for self-correction in the face of new conditions and concrete evidence. Defending it is not merely an epistemological matter but an ethical one, in which commitment to truth as method must guide human endeavour.

**Keywords:** common sense; *sensus communis*; Reid-Moore tradition; artificial intelligence; large language models; social epistemology; epistemic commons; algorithmic motivated reasoning; cognitive colonialism; attention economy; information ethics; digital governance; cultural hegemony; social technology.

---

## Introduction

You have most likely witnessed scenes like these yourself:

Rain is falling. The sound of drops against the windowpane fills the room. A man sits on the sofa, phone in hand, wanting to know whether the weather is good enough for a walk outside. Rather than glancing at the window, he checks a weather app. Does he trust the algorithm more than his own senses?

After a long and busy day, you notice a racing heartbeat, a headache, or pain in your shoulder and arm. You immediately enter these symptoms into a search engine. Within seconds, a flood of contradictory information and alarming diagnoses appears before you, ranging from brain tumours to imminent cardiac arrest. You panic, and search further. It does not occur to you that you have spent the whole day talking to people, working hard, and drinking too little water. You set aside everything you know about your own condition and the most basic principles of self-care, and tie your health to the advice of unqualified online sources.

At a family gathering, a debate breaks out over a scientific or social issue. Everyone immediately reaches for their phones to extract the "truth" from search engines, artificial intelligence, or the algorithmic feeds of social media. Has the social infrastructure of shared reality collapsed? Do we no longer trust our own reasoning and judgement? Or have artificial intelligence and Google simply taken the place of human experience and common sense?

Imagine a citizen on the eve of a decisive election or a major economic decision. Rather than analysing positions and assessing realities objectively, he surrenders his mind entirely to the stream of data recommended by platforms. He believes he is deciding freely. In reality, his political and economic preferences are being engineered in an invisible factory of manufactured consent, by algorithms optimised for anger and engagement. Has the political common sense that should serve as the first line of defence against demagoguery been replaced by machine-made motivated reasoning and algorithmic manipulation?

These scenes, and countless others we observe in ourselves and in others every day, open a window onto a bitter reality of the twenty-first century: we have surrendered our accumulated knowledge, the experience of millennia, our human reason and judgement, and even our direct sensory grasp of the world, to machines. In the simplest matters of daily life, we have become dependent on artificial intelligence and algorithms. This is a situation without precedent in history.

By "common sense" in this article, we mean what human beings have relied upon for thousands of years to resolve personal, familial, and social problems: that inner sense and shared understanding, often unspoken yet held in common, which told people that fire burns, that jumping from a height is fatal, and that dishonesty and betrayal are wrong. It is the body of self-evident and shared beliefs by which people lived without explicit reasoning, the hidden pillar of everyday life.

This understanding of "common sense" operates across three interconnected layers: the cognitive layer (the basic beliefs on which each person acts without explicit reasoning), the social layer (the shared infrastructure of reality that binds communities together), and the normative layer (the commitment to truth as method). Common sense, on this account, is not infallible. It is fallible, susceptible to deception, and can fail to read a situation correctly. What distinguishes it from the machine reasoning we see today, however, is that human common sense grows from lived experience, registers the concrete and empirical consequences of decisions and interpretations of reality, and retains the capacity to return and correct itself when confronted with strong and clear evidence.

If we wish to trace this concept through history, it appears that for thousands of years common sense had no special name, yet it served, unspoken and unwritten, as the standard of right living for ordinary people in their daily affairs. Perhaps the concepts that ancient peoples developed in

their cultures as simple and direct guides to life were pointing towards precisely this: concepts such as Ma'at in ancient Egypt, Dharma for the Hindus, Kherad in Zoroastrian thought, and Dao in ancient China. Although the roots and origins of each of these practical guides may have drawn on nature, ancestors, or gods, the functional meaning and purpose behind all of them was the same: a set of practical self-evident truths for living rightly, an order that everyone knew and in whose correctness no one doubted. This shared wisdom was neither theory nor argument. It was a guide to survival and equilibrium in material life.

With the Greeks, the scene changed. Socrates, Plato, and Aristotle were the first to lift common sense from the practical plane to the epistemological, asking: if the self-evident truths of every society differ, what is the genuine standard of knowledge?

Aristotle introduced the concept of "*koinē aīsthēsis*" (common sense): a perceptual faculty that integrates the data of the various senses. By emphasising experience and logic, he brought knowledge down from the heaven of Platonic Forms to the ground of human experience.

This meaning later acquired a more social and political character in Rome, particularly in Cicero, who called it "*sensus communis*". By extending the concept from individual judgement to collective conscience, and linking it to the idea of the *consensus gentium*, common sense became a social and political concept that simultaneously claimed universality.

In the Middle Ages, Thomas Aquinas wove this inheritance together with Christian theology, connecting it to the "first principles of morality" and to the preamble of knowledge of God. Up to this point, common sense remained a largely unified and trustworthy concept, bound to metaphysics.

With the Renaissance came a turning point. Common sense was, for the first time, separated from metaphysics and brought down to the ground of real life. Lorenzo Valla sought common sense not in philosophy but in the "language of the people" and in "historical consensus" (*consensus omnium*). He believed that what people had collectively regarded as true throughout history was the standard by which claims to power could be measured. Using precisely this linguistic tool, he demonstrated that the Donation of Constantine (upon which the Church rested its claim to ownership of half of Europe) was a forgery, because its language was inconsistent with the usage of Constantine's era. Valla showed that common sense is constructed from the history of lived language, not from philosophical argument, and it is this quality that makes it an instrument of critique against power (Rosenfeld, 2011).

In the fifteenth century, Erasmus extended the concept to "*sensus communis humanitatis*": a universal sense that could serve as the foundation for understanding and peace among nations. He advanced the ideal of universal access to knowledge, an ideal that the invention of printing in the 1450s brought from dream to reality.

Montaigne, in the sixteenth century and in the midst of the French Wars of Religion (1562-1598), called everyday common sense "the stronghold of humanity against fanaticism and absolute certainty". In this period, truth was no longer the monopoly of philosophers and priests. It had taken root in the shared judgement of ordinary people.

In 1637, however, Descartes, in his slim volume "Discourse on the Method", swept away this prevailing confidence in the value of common sense and opened a great rift in human thought. Seeking an unshakeable certainty, he declared the shared experience of humanity unreliable and deceptive, demolishing everything except "I think, therefore I am" (Descartes, 1637/2006). With this move, Descartes drove a wedge into the concept of common sense, pulling it down from its seat of authority and demoting it from a civic virtue to a source of error and prejudice. The result, which we may call the "Cartesian rupture in common sense", was the fracturing of European

thought. The Continental rationalists (Spinoza, Leibniz) sought certainty in pure mathematical reason and rejected common sense, whilst the British empiricists (Locke, Berkeley, Hume) sought to rescue it by grounding it in human sensory experience.

The debate continued, and in Naples in 1725 another thinker opened a third way. Giambattista Vico stood against Descartes and argued, through the principle "*verum ipsum factum*" (the true is what is made), that human beings genuinely know only what they themselves have made: language, law, myth, institutions, and history (Vico, 1725/1948). From this perspective, common sense was neither a source of error nor merely individual sensory data. It was the "collective and historical wisdom of a people", something living and latent in shared language, proverbs, and customs. Vico was ignored in his own time, but his idea of common sense as a collective and cultural matter became one of the most fertile branches of modern thought in the nineteenth and twentieth centuries, and connects directly to contemporary critiques of cultural bias in artificial intelligence.

The eighteenth century may be called the last golden age of common sense, a century in which one could still speak with confidence of the "shared reason of all humanity". In this same century, both the greatest threat and the strongest defences emerged. Hume shook the foundations of common sense by reducing basic beliefs to mere "habit" (Hume, 1748/2007). Thomas Reid and the Scottish School of Common Sense responded by enumerating the "first principles" of common sense and placing the burden of proof on the sceptic (Reid, 1788/2010). Kant simultaneously rescued common sense from Humean scepticism and confined it, declaring it valid but only within the "phenomenal world" (Kant, 1781/1998). In this same century, common sense made its way from the philosopher's study to the coffeehouses, the newspapers, and eventually to the arena of the French Revolution. When "general consensus" became "general will", the blade of the guillotine demonstrated what horrors could be wrought by the claim to represent the absolute "common good".

With the arrival of the nineteenth century, the optimism of the eighteenth century regarding common sense and popular belief came to an end. Thinkers from Burke and Herder to the Romantics and Hegel showed that the content of common sense (what appears "self-evident" in any given society) is deeply historical, linguistic, and cultural. What seems reasonable in one era may appear absurd in another. Common sense was no longer the simple, fixed, and universal capacity of the Enlightenment. It had become a fragile and context-dependent phenomenon. The lasting lesson of the nineteenth century was that common sense must be treated like a garden: it requires constant care, education, institutional support, and critique.

The twentieth century took this fragility to its extreme. With the outbreak of the First World War, the bitter question arose: if common sense could not prevent such destruction, what use was it? In response to this devastating challenge, George Edward Moore, in 1925, defended common sense by defending the "self-evident" (Moore, 1925). The later Wittgenstein redefined it as a "form of life" and the silent backdrop of all meaning (Wittgenstein, 1953). The Scottish tradition continued and was revived in the work of thinkers such as Chisholm and Plantinga.

By the end of the twentieth century, common sense had survived, but it was no longer innocent or unified.

Today, in the first quarter of the twenty-first century, this ancient concept faces a threat of a new kind. Throughout that long history, one assumption held constant: common sense is present in the mind of every healthy person and is shared by all. Today, that assumption is collapsing. We are daily submerged in a flood of claims and counter-claims, uncertain which doctor, politician, or advertisement to trust. Worse still, with the arrival of artificial intelligence and its capacity to produce entirely fabricated audio and imagery, we can no longer trust even what we see with our

own eyes. The fundamental difference between today's crisis and all previous ones lies precisely here. Common sense has not died, but for the first time in history, its production, distribution, and control have passed out of human minds and cultures and been transferred to the algorithms of social media platforms, large language models, and the policies of states.

These momentous changes have transformed the question we must ask about common sense. Today, in the age of algorithms, the three epistemological layers we defined at the outset of this introduction (the cognitive, social, and normative layers) have become so deeply intertwined that any attempt to separate them artificially is a simplification. The crisis of common sense today lies precisely in this entanglement. Where once we asked what common sense is, or how long it will survive, we must now ask who constructs it, at what scale it is distributed, and whether its nature remains human. In this sense, the twenty-first century may be called the era of "the war over common sense".

This article addresses that war in three steps: first, an examination of the four main intellectual currents of this quarter-century; then, an analysis of the three real fields of struggle (ownership, territory, and identity); and finally, an answer to the question of why we believe common sense remains, in our time, the last bastion.

The article ultimately argues that, amid this upheaval, common sense (with its rootedness in shared experience and its capacity for self-correction in the face of evidence) may be our last bastion against machines that can justify anything through apparently logical reasoning. Defending it is not merely an epistemological matter but an ethical one. To understand this war, common sense must be seen for what it truly is: not a silent self-evidence in the mind of every healthy person, but a "social technology", one that can be designed, dismantled, and rebuilt.

---

## **Part A: Intellectual Currents Concerning Common Sense in the Twenty-First Century**

In the opening quarter of the twenty-first century, four main currents have taken shape around the concept of common sense:

- analytic epistemological defence (2000 to 2015);
- cognitive engineering of common sense and the construction of an artificial version (from the early 2010s to the present);
- the crisis of the shared social infrastructure of common reality (2011 to the present);
- and nascent efforts to rebuild its digital infrastructure (from around 2015).

These currents have developed independently and in parallel, with occasional overlap, but they share a common message: common sense has not died, yet its production, distribution, and control are no longer the exclusive preserve of the human mind and human culture.

### **Current One: The Analytic Epistemological Defence of Common Sense**

From the beginning of the 2000s, and particularly in the first half of the 2010s, a new wave took shape within analytic epistemology. Its aim was to find a rational basis for defending the basic and self-evident beliefs of everyday life: beliefs such as the existence of the external world or the reality of ordinary objects, upon which people act without explicit reasoning. This current continues a tradition rooted in Thomas Reid (eighteenth century) and George Edward Moore (early twentieth century), the Reid-Moore tradition, which held that some beliefs are so fundamental that the burden of argument falls on the sceptic, not on the believer.

The central question of this current is simple but important: how can entirely self-evident beliefs be kept rational and defensible against sceptics who doubt those very certainties?

The philosophers of this current seek an answer that neither entangles them in an endless chain of justifications nor compels them to find an indubitable Cartesian foundation. This is precisely the difficulty the Reid-Moore tradition has wrestled with from the outset. Three main approaches, each a sub-branch of this current, have emerged over the past two decades.

**1. Phenomenal conservatism.** Drawing on Michael Huemer, who holds that "if something seems true, that appearance is the starting point of justification" (Huemer, 2007): if something appears true and we have no reason against it, accepting it is warranted. In the 2020s, however, critiques emerged. The spread of digital representations and, in particular, the phenomenon of deepfakes showed that "apparent truth" is not always a reliable guide. Revised versions were therefore proposed that place greater emphasis on context, the type of experience, and the likelihood of perceptual distortion.

**2. Perceptual foundationalism and contemporary dogmatism.** This sub-branch draws on the work of James Pryor, in which "ordinary perceptual experience" provides *prima facie* justification on its own, without the need to refute competing hypotheses (Pryor, 2000). The mere appearance that "there seems to be a hand here" is sufficient to begin justification. The contested point in this view is the definition of "normal conditions". In a world saturated with virtual reality and digital manipulation, determining which experiences are "ordinary" and "undeceived" is no longer straightforward. Most objections to this view target precisely these practical boundaries, not its logical structure.

**3. Contemporary re-readings of the Reid-Moore tradition.** Philosophers such as Scott Sturgeon, Declan Smithies, and Maria Lasonen-Aarnio have sought to connect the classical defence of basic beliefs to newer theoretical tools (Sturgeon, 2020; Smithies, 2019; Lasonen-Aarnio, 2014). The result is a more coherent philosophical framework, but most of these discussions have remained at the conceptual level, engaging less with the problems of the contemporary world: problems such as the collapse of shared common-sense norms in polarised online societies. This gap between the individual dimension of defending basic beliefs and social realities is the practical weakness of this approach, at a point when shared certainties are no longer stable and the crisis of "online testimony" undermines individual justification.

After a quieter period (2015 to 2018), thinkers within this approach launched a new wave from around 2019, this time focusing on mediated experiences, technology-induced perceptual errors, and the implications of artificial intelligence for basic beliefs.

In summary, the first current has succeeded in showing that many everyday beliefs require no certain foundations. It has not yet answered the question of why those same beliefs, once widely shared, have become so contested and unstable in an environment where platforms and networks constantly reshape the cognitive context. This gap opens the way for the other currents.

## **Current Two: Understanding Human Cognition and the Dream of Building Artificial Common Sense**

Against those who regard common sense as an ancient inheritance or a pure self-evidence, a group took a different approach. They proposed that common sense is neither a fixed essence nor a mere instinct, but a pattern of cognitive behaviours that can be studied, modelled, and to some extent reproduced, whether in the brain or in the machine.

The background to this current reaches back to the 1950s, when some psychologists and cognitive scientists sought to show that human common sense consists of a set of adaptive mechanisms.

Herbert Simon's theory of "bounded rationality" (Simon, 1957), and the research that followed, demonstrated that people make reasonably good decisions with limited cognitive resources.

Along the same lines, Gerd Gigerenzer, in his theory of "ecological rationality" (2007), showed that many of our apparently simple judgements result from adaptation to real environments, not from the application of abstract rules (Gigerenzer, 2007). Cultural and anthropological studies have further shown that part of common sense resides not in the individual brain but in institutions, customs, language, and social cooperation: what Joseph Henrich calls "cultural accumulation" (Henrich, 2020).

The combined result of this research was a multi-layered view (cognitive, ecological, and cultural) that transformed common sense from something "individual and private" into a distributed, shared, and contextual phenomenon. From within this multi-layered view, an ambitious idea was born: if we understand the mechanisms of human common sense, perhaps we can build a version of it in a machine. This idea developed along three paths.

**1. The symbolic and neurosymbolic approach.** This involved equipping machines with rules, world knowledge, and formal reasoning tools. It sought transparency and control but repeatedly ran up against the limits of a world full of ambiguity and unpredictability.

**2. The statistical approach and large language models.** Models trained on vast quantities of text to extract linguistic patterns became, particularly after 2023, considerably more capable of language comprehension and the production of "common-sense-like responses". Even so, the absence of tacit knowledge, stable working memory, a world model (a structured representation of physical reality and causal relations), and weakness in causal reasoning remain the principal limitations.

**3. The embodied and interaction-based approach.** This group argues that common sense cannot develop without a body, sensory experience, and real engagement with the world. From this perspective, purely textual models, however large, face a structural ceiling. They cannot acquire self-evident physical knowledge, deep causal reasoning, or sensorimotor understanding, because these are inherently dependent on embodied experience (Dreyfus, 1972).

Hubert Dreyfus had made this argument earlier and with greater force: symbolic artificial intelligence was doomed from the outset, because human common sense springs not from abstract rules but from "being-in-the-world", from the body, from situatedness, and from direct engagement with physical and social reality (Dreyfus, 1972). Large language models reproduce this structural limitation in a new form: the volume of data cannot substitute for a body.

The current state of this field shows considerable progress, but a large gap remains between that progress and the goals originally set. Models have improved on simple inferences, understanding everyday contexts, and some standard common-sense criteria. The research consensus, however, is that the tacit knowledge of human beings (what comes from experience, embodiment, and culture) has not yet been reproduced in a machine. These systems are weak in causal reasoning, understanding intention, maintaining a stable identity, and modelling the world. Researchers in this field believe that many of the "common-sense-like" behaviours of these models are products of linguistic pattern-matching, not genuine understanding. The prospect of an "artificial version of common sense" therefore remains largely experimental, and the evidence does not suggest that it can be reached simply by adding more textual data or increasing scale.

The important achievement of this current, nonetheless, has been to deepen and broaden our understanding of human common sense: from something self-evident and silent, to a subject amenable to study and critique, with cognitive, cultural, and social roots rather than a fixed "essence".

### **Current Three: The Collapse of the Shared Social Infrastructure of Common Sense**

This current extends a long-standing view (associated with Bourdieu, Foucault, and Gramsci, among others) that common sense has always been a battleground of power. Researchers in this field, however, argue that with the dominance of social networks (from 2011 onwards) and the rise of artificial intelligence, the tools of that contest have been transformed entirely. In response to the central question of why contemporary societies cannot reach agreement even on the most basic facts, this current answers that the problem lies not in individual cognitive failure but in the loss of a shared basis for reality. This collapse has three layers: structural historical pressure, institutional erosion, and platform architecture.

**1. Structural pressure and the power struggle over common sense.** Before algorithms and social networks entered the field, a long tradition in critical thought held that common sense has always been a battleground of power. What appears "self-evident" and "natural" in any society is often the product of power relations that have concealed themselves. Part of today's collapse is not accidental but the result of an organised project to manufacture doubt and destabilise shared reality.

Pierre Bourdieu named this phenomenon "doxa": a body of beliefs and values so thoroughly internalised that they are experienced not as a "point of view" but as "reality" itself, and are for that very reason the most powerful instruments of domination, because they are invisible (Bourdieu, 1977).

Michel Foucault extended this idea through the concept of "régimes de vérité": every period has its "dispositifs" (arrangements or mechanisms) that determine what counts as "reasonable" and what as "madness", who is "expert" and who is "ignorant". These mechanisms are never neutral (Foucault, 1980).

Friedrich Hayek, sharing the same concern but from a wholly different angle, warned that the real danger lies precisely in a central power seeking to define and impose common sense from above. In his view, genuine common sense wells up from below, from the millions of dispersed decisions of free individuals: what he called "dispersed knowledge", which no central planner has the capacity to gather (Hayek, 1945).

Frantz Fanon showed from yet another angle that "universal common sense" is often nothing more than European colonial common sense, imposed on the rest of the world by force and culture, and that decolonisation must above all be a decolonisation of the mind (Fanon, 1961).

In the same tradition, though with differing emphases, one may name Gramsci and his cultural hegemony, Althusser and his ideological state apparatuses, Proctor and the deliberate production of ignorance, Fricker and epistemic injustice, Harding and standpoint epistemology, and Mignolo and cognitive colonialism. Each illuminates a different facet of the same reality.

**2. The erosion of truth-producing institutions.** Shared common sense is the product of institutional coordination (reputable news agencies, universities, judicial and scientific systems) that once stabilised the "baseline of reality". Jonathan Rauch (2021) likened this network to society's "epistemic operating system". The erosion of these institutions' credibility, whether through organised attack or internal weakness, blurs the boundary between "consensual reality" and "the narratives of echo chambers" (Rauch, 2021). In short, shared self-evident beliefs cannot remain stable without institutional maintenance.

Hannah Arendt had earlier shown that public life requires a "shared reality" for its democratic functioning, not agreement on values, but agreement on facts. When this shared ground collapses, not only dialogue but the very possibility of politics disappears (Arendt, 1967).



**3. Platform architecture and the attention economy.** Research conducted after 2011, and particularly after 2016, has shown that algorithms, by privileging emotionally charged and polarising content, drive users into epistemic bubbles. This is not a conspiracy. It is the straightforward result of the attention economy: optimisation for engagement, not for truth.

A much-cited and contested example is YouTube's recommendation algorithm in the mid-2010s. Some studies found that it gradually steered users towards more extreme content. Other studies challenged this finding, arguing that the algorithm amplifies existing demand rather than producing radicalisation directly. What remains constant across both accounts is the same underlying reality: these algorithms are optimised for engagement, not for truth.

These three layers reinforce one another. Structural historical pressure prepares the ground; institutional erosion breaks down authority; and algorithms fill the resulting vacuum with polarising content. Michael Lynch (2016) shows how this cycle undermines the epistemic trust of citizens, to the point where people no longer know from whom to obtain reliable expertise (Lynch, 2016).

Neil Levy (2021) attributes this situation to the "pollution of the epistemic environment". Dan Williams (2023), however, locates the root cause in identity and coalitional motivations: groups that choose a version of reality not on the basis of evidence but on the basis of political or cultural identity (Levy, 2021; Williams, 2023).

The most striking example of this was seen during COVID-19: two "independent epistemic communities" that could not even agree on whether the disease posed a genuine danger. Levy argues that in a "poisoned epistemic environment" one cannot expect people to reliably arrive at correct beliefs. Our moral expectations of individuals' epistemic choices must therefore be adjusted to the reality of that environment (Levy, 2021). In his view, cultivating individual epistemic virtues is necessary but insufficient against the pressure of destructive structures. Responsible policy is also required.

C. Thi Nguyen draws a subtle distinction between an "epistemic bubble" (in which contrary information is simply filtered out) and an "echo chamber" (in which trust in contrary sources is actively destroyed). The second is more dangerous, because it eliminates not information but the very capacity to receive information (Nguyen, 2020).

This crisis does not take the same form everywhere. In the West, the explosion of sources has produced a kind of "extreme epistemic pluralism" in which everyone has their own reality. In countries such as China and Russia, state control of information has created an official and exclusive narrative. In both cases the result is a weakening of shared common sense, but in two different forms: in the West, through chaos and uncontrollable proliferation; in the East, through enforced uniformity.

We also face a newer crisis: deepfakes and the collapse of the concept of witnessing. Research shows that today's crisis results from the simultaneous abundance of suspect information and scarcity of reliable information. Naffi (2025) warns that if detection tools are not strengthened, "trustworthy reality" will collapse rapidly. Throughout history, "seeing with one's own eyes" was the standard of credibility. In the world of deepfakes, can we still trust what we see and hear? Research has shown that manufacturing doubt and destabilising shared reality has at times become a deliberate political strategy.

Lewandowsky and colleagues have shown that the problem is not simply "wrong data" but the quality of citizens' entire "information environment". The issue is the quality of citizens' "cognitive diet": just as an unhealthy diet poisons the body, an unhealthy information diet poisons the mind (Lewandowsky et al., 2017).

Taken together, these perspectives present a large picture of the algorithmic world. The algorithms of a handful of large corporations simultaneously intensify all three layers: Bourdieu's invisible doxa (imposing self-evident beliefs that no one experiences as "imposed"), Foucault's régime of truth (determining which voices are heard as "reasonable"), the concentration of power that Hayek feared (a few technology giants in place of millions of dispersed decisions), and the cognitive colonialism that Fanon described (carried out this time not with cannon and rifle but with data and algorithms).

In this way, common sense today is under siege, not from philosophical scepticism (which has always existed), but from structural historical pressure, institutional erosion, and platform architecture, each reinforcing the others. This siege has brought about the loss of that "epistemic commons" that makes constructive coexistence and mutual engagement possible. Hess and Ostrom introduced this term in 2007 in the political economy of knowledge to describe "information as a shared resource" (Hess and Ostrom, 2007). It refers to that rich and common ground of reality, trust, and language: a space as invisible as air, yet without which intellectual breathing is impossible. From this perspective, shared common sense is precisely a public good. It belongs to no one, yet everyone depends upon it and lives within it. Like the natural environment, its degradation is slow and imperceptible, but its consequences are catastrophic.

Until this "epistemic commons" or "shared infrastructure of reality" is rebuilt, any discussion of common sense (human or artificial) rests on unstable ground. The consequences of this siege are not only epistemological but political: democracy cannot function without shared common sense. As Sunstein showed in "*#Republic: Divided Democracy in the Age of Social Media*" (2017), democracy requires a minimum of shared reality if citizens are to reason together rather than merely stand in opposition to one another. The collapse of that shared reality is the collapse of the very possibility of democratic dialogue (Sunstein, 2017).

#### **Current Four: Nascent Efforts to Rebuild the Digital Infrastructure of Common Sense**

Where the third current showed the loss of the "epistemic commons", the fourth seeks to rebuild it. Unlike the three preceding currents, this one, even in mid-2026, does not yet constitute a coherent "school". It began around 2015 and remains scattered and multi-rooted. What brings these (for now) dispersed efforts together is a single shared goal: to rebuild the shared infrastructure of reality in a world whose information environment has collapsed.

Broadly, these efforts combine platform engineering, information ethics, participatory design, institutional architecture, and digital policy. They seek answers to the two crises of the third current (institutional weakness and polarising architecture) along two axes: redesigning platforms on the basis of human values, and intelligent governance. This current holds that "design without governance" leads to a chaotic marketplace, and "governance without design" to an ineffective bureaucracy. The two must therefore advance simultaneously. Two paths can be identified within this fourth current.

**Path one** locates the root of the crisis not in "the ignorance of the people" but in the poor design of the information environment. Luciano Floridi argues that the informational quality of the environment is a public good, like clean water. Just as industries have no right to pollute the natural environment, technology companies cannot, without accountability, poison the epistemic ecosystem (Floridi, 2013). This framework now underpins many of the European Union's policies.

Among the contributors to this path, Shannon Vallor emphasises "technological virtues". She holds that honesty in engaging with data, epistemic humility, caution, empathy with other perspectives, and curiosity in the pursuit of truth are the five prerequisites of shared common sense in an environment that constantly pushes us towards confirming our existing assumptions (Vallor, 2016).

A practical and successful example from this path is the "Community Notes" project on the platform X. Under this scheme, a corrective note on a piece of content is only displayed when users with differing orientations agree that it is helpful: a sign that political disagreement need not prevent epistemic cooperation, provided the platform and its conversational architecture are well designed.

**Path two** asks: how can law and oversight be used to build "spaces resistant to manipulation"? Hélène Landemore argues that digital democracy can activate "collective wisdom". Tools such as Polis in Taiwan have shown that, with the right conversational architecture, even polarised societies can reach broad agreements (Landemore, 2020).

Another institutional example is the European Union's legislative package comprising the Digital Services Act (DSA) and the AI Act, which targets algorithmic transparency, the reduction of systemic risks, and the content responsibility of large platforms. A concrete instance of this framework in action was the 405-million-euro fine imposed by Ireland's Data Protection Commission against Meta in 2022 under the GDPR, for violating children's data privacy on Instagram.

Researchers drawing on Habermas have also sought to reconstruct the concept of a "digital public sphere": a space that guarantees civil dialogue and independence from both state and private power (Habermas, 1989).

These programmes extend beyond the West. Perhaps the most significant is India's "Digital Public Infrastructure" (DPI) strategic framework. This model is neither purely regulatory in the European style nor a laissez-faire American approach. Its aim is to build public infrastructures on which the private sector competes. India has so far concluded cooperation agreements with around 23 countries, and its identity platform (MOSIP) is in use in more than 20. The "50 in 5" campaign aims to bring fifty countries on board, though a significant difference exists between "interested" and "adopted". The implementation of this path also carries risks of state surveillance, the concentration of power, and violations of civil liberties and rights.

In summary, the fourth current has yet to take a final coherent shape, but its achievements are tangible: governance legislation in Europe, participatory tools in Taiwan, and infrastructure design in India. Its contributors have shown that good design can construct "shared reality". The limitations, however, are serious. Most solutions have been tested only at limited scale; the cost of regulation is high; the risk of censorship accompanies every act of regulation; cultural differences obstruct global replication; and large corporations possess enormous resources for circumventing rules. Perhaps the intellectual foundation of all these efforts rests on Floridi's "information ethics": the last shared framework that can serve as the basis for transnational dialogue about technology and society.

**Summary of Part A.** In the twenty-first century, common sense is no longer a purely philosophical or individual category. It is a network of basic beliefs, social structures, and digital infrastructures. The first current emphasises the philosophical defence of everyday beliefs. The second shows that human common sense is ecological, distributed, and to some degree simulable. The third highlights the crisis of the shared infrastructure of reality. The fourth presents practical efforts to rebuild it, through ethics, design, and governance. The key conclusion is that common sense is still alive, but its centre of gravity has shifted from the individual mind to social and digital environments. The principal challenge ahead is the design and oversight of these new infrastructures, with due regard for values, the protection of individual rights and freedoms, and justice.

Taken together, the picture these four currents present is a reminder that common sense is not an innate individual property but a social technology: a tool that societies have built to reduce

cognitive costs, enable collective coordination, and allow people to live alongside one another. Like any technology, it can be designed, dismantled, or rebuilt. And this is precisely what is happening today, before our eyes.

---

## **Part B: Three Battles over Common Sense in the Twenty-First Century**

From within those four currents, one large reality emerges. Contrary to previous centuries, the question is no longer how far common sense is dead or alive, reliable or fallible. The question is who constructs it and owns it, at what scale it is distributed, and whether it will ultimately retain a human character. The common sense we once sought in the mind of every healthy person is today passed between servers, algorithms, implants, and global institutions. These three great questions have drawn us into three fields of battle: the battle over ownership, the battle over territory, and the battle over identity. These three fields are not of the same kind. In one the issue is power, in another scale, and in another the very nature of what common sense is. This reflects the character of common sense in the twenty-first century: a subject that is itself multi-layered and resists any single-dimensional framing.

### **1. The Battle over Ownership**

Who owns our common sense? Are "we" still its bearers, or has something else taken our place? From the eighteenth century, the belief held that common sense lived in the mind of every healthy person. In the nineteenth century it came to belong to the nation and the class. In the twentieth, ideologies and sciences appropriated it. In the twenty-first century, ownership has passed to three new actors, none of whom is fully human: corporations (Meta, Google, OpenAI), which own the algorithms and the data; states (China, the European Union, the United States), which own the laws and the filters; and machines (language models and automated systems), which manage the production and reproduction of self-evident beliefs.

Ownership here does not mean ownership in the legal sense. It means control over the process by which self-evident beliefs are formed: who determines which belief appears "natural", which "reasonable", and which "delusion".

Broadly, these three large actors contest this most intensely in two fields: social networks and artificial intelligence.

**1.1. Social networks.** With the rise of the internet and the attention economy (particularly from 2006 onwards), common sense is no longer the accumulated lived experience of human beings and the product of face-to-face conversation. It is constructed by the algorithmic feed. Eli Pariser (2011) showed that each user now lives within a personalised version of self-evident beliefs, so that what is self-evident to me may appear delusional to my neighbour (Pariser, 2011). As noted in the preceding section, Sunstein warns that democracy cannot function without a shared sense of reality, and that algorithms have destroyed precisely that sense (Sunstein, 2017).

Shoshana Zuboff, in "The Age of Surveillance Capitalism" (2019), argues that the attention economy has inverted the values of common sense: whatever provokes the greatest anger, fear, or affirmation comes to appear the most "shared" (Zuboff, 2019). In the words of Tom Nichols in "The Death of Expertise" (2017), algorithms now play the role of a kind of "council of common sense", determining which belief appears "self-evident" and which a "conspiracy theory" (Nichols, 2017).

The darkest horizon is opened by Nick Bostrom (2014). Where Zuboff showed how corporations take ownership of our data, Bostrom raises the possibility of a situation in which the very process by which common sense is formed becomes vulnerable to a "superintelligence": a point at which

the question is no longer ownership of data but ownership of human "judgemental capacity" and "cognitive agency". From this perspective, the battle over ownership is not merely an economic competition. In an uncertain future, it becomes a question of the survival of human intellectual independence. When machine-generated common sense is no longer a neutral public resource but a commodity to be extracted and sold, we are very far from the Ciceronian *sensus communis* that sought consensus in public understanding and the building of a "higher civic wisdom".

**1.2. Artificial intelligence.** The second field involves large language models, which have themselves become producers of common sense. For the first time in history, a non-human entity can not only distribute common sense but produce it. Since 2022, models such as ChatGPT, Claude, and Gemini have given responses that users often regard as self-evident, even when those responses are fabricated.

These systems are no longer merely "tools". They are well on their way to becoming the new arbiters of what is taken for granted, a development still in its early stages but accelerating at a worrying rate. Perhaps it is better to call them "thinking machines" rather than artificial intelligence.

There are growing signs that in many societies young people, when asking "is this right?", turn increasingly not to grandparents, teachers, or parents, but to chatbots. The chatbot draws its answer from "linguistic co-occurrence patterns" across billions of texts: an approximate reflection of expressive patterns, not a direct representation of common sense.

The effort to build "artificial common sense" is not new: from the Cyc project in the late 1980s, to Open Mind Common Sense and ConceptNet in the late 1990s, and then to ATOMIC and COMET, which combined neural networks with knowledge bases as deep learning emerged in the late 2010s. The message of these efforts is clear and sobering: common sense is no longer lived human experience but an engineered product that can be produced, updated, censored, or optimised. If this trend continues, a large part of the production of self-evident beliefs will be delegated to technological institutions. Optimism about there being "no clear winner" in this battle is misleading, because we no longer set the rules of the game, and the absence of the power to set rules is, in practice, equivalent to losing.

In the battle over ownership, the immediate and concrete danger is this: when the definition of "self-evident" is transferred from the human mind to corporate servers, intellectual independence is no longer a natural right. It becomes a privilege that must be rented from the algorithm.

## **2. The Battle over Territory**

Can common sense become global, or will the attempt to make it planetary lead to its collapse? Today, for the first time, a single version of "what is self-evident" is simultaneously available to eight billion human beings, not through books and newspapers but through language models and algorithms that declare what is "reasonable" and "natural". This is the greatest leap in the scale of common sense in history, and it is potentially as dangerous as it is unprecedented.

To grasp the scale of this danger, we must return to a biological constraint. According to "Dunbar's number", the human brain evolved for life in small groups and can sustain stable relationships with no more than approximately 150 individuals, not with millions (Dunbar, 1992). Yet large language models simultaneously answer billions of people and define the framework of their self-evident beliefs. Common sense is no longer built from the bottom up (from local experience to national consensus) but imposed from the top down by a handful of global models onto the minds of each one of us.

This novel phenomenon raises several critical questions. Does global homogenisation destroy cultural diversity and local wisdom? Will people trust a system that defines their self-evident beliefs? And rather than a single global common sense, will we witness fragmentation and competition between American, Chinese, and European "common senses", each carrying its own biases?

The question grows heavier when we attend to a bitter reality: at the very moment we fear algorithmic homogenisation, we are also in urgent need of a form of "planetary thinking". Climate change, water crises, pandemics, and climate-driven migration are "species-level" crises that cannot be resolved by local common sense. Tim Lenton and Bruno Latour, in their article "Gaia 2.0", argue that humanity can for the first time intervene consciously in planetary processes, and that this awareness must lead to a form of common sense that regards the whole Earth as the primary actor (Lenton and Latour, 2018; Latour and Weibel, 2020).

Donna Haraway goes further, calling for a future common sense so "multi-species" that it treats the interests of forests and microbes as self-evident (Haraway, 2016).

Here, however, lies a fundamental paradox. The planetary common sense that Latour and Haraway seek is a "democratic systemic awareness": simultaneous understanding of shared problems and respect for diversity and indigenous knowledge, built from the bottom up. What is actually happening is top-down algorithmic homogenisation, defined by a handful of corporations and states, which dismisses cultural diversity as "inefficiency" and indigenous knowledge as "unreasonable", and which is built upon data that itself reflects existing power and inequality.

All of this unfolds whilst our brains still operate at the scale of a village from five or ten thousand years ago, telling us "what I see with my eyes, I believe". When science speaks of the melting of the polar ice caps, that ancient common sense replies: "the weather has always changed; I see no difference myself." The result is a strategic paradox: local and ancient wisdom is insufficient for understanding planetary crises, yet the solution cannot be the imposition of a single algorithm.

This, of course, is an instance of a limitation in common sense, not a reason to reject it. A properly functioning common sense ought to yield to the accumulated evidence of climate science. If it does not, this is a sign that something (ideology, self-interest, or an information bubble) has erected a wall between common sense and the evidence, not that the nature of common sense is inherently resistant to reality.

From the examination of the battle over territory, at least two things become clear. Common sense in its local and traditional form is insufficient, by itself, for the survival of the species. And the planetary common sense that today's algorithms are constructing is neither democratic nor just. We therefore need deliberate and political intervention, not to resist planetary common sense, but to determine the path of its formation: a common sense built from the bottom up through the dialogue of communities, respectful of diversity, transparent, and democratic. Otherwise, not only the climate but trust, democracy, and humanity itself are at risk.

### **3. The Battle over Identity**

The question of whether, in 2050, the character of common sense will still be "human" has made the battle over identity the most important of the three. Will common sense continue to spring from lived experience, the body, emotion, and human dialogue, or will it become an algorithmic and statistical process?

Even today, when a young person asks ChatGPT "is it right to...?", on whose statistical consensus does the answer rest? Henrich and colleagues showed in 2010 that the findings of psychology are largely built on societies described as WEIRD (Western, Educated, Industrialised, Rich,

Democratic), and that these societies are not psychologically representative of humanity (Henrich et al., 2010). This same critique applies today with even greater force to language models trained predominantly on English and Western data: what might be called "the silent rewriting of Western cultural self-evident beliefs". This is a quiet, gradual, and apparently "helpful" process (because the answers are usually correct), a form of cognitive colonialism without a coloniser and without violence, carried out through data and algorithms. The result is that the concerns of a young person in Jakarta or Tehran gradually come to resemble those of a young person in California or London, not through dialogue but because of the "weight of the data".

Yuval Noah Harari, in "Homo Deus", constructed, through his account of the merger of biology and algorithm, a vision of a future in which human identity stands on the threshold of redefinition: a vision that has proved influential for popular understanding of the human future (Harari, 2016).

Miranda Fricker calls this process "epistemic imperialism": the structural exclusion of groups from the tools for interpreting their own experience, which is precisely what Western-origin language models do through their training data (Fricker, 2007).

Emily Bender and colleagues, in the well-known article "Stochastic Parrots" (2021), showed that large language models do not "understand" but produce "statistical reproduction of linguistic patterns". Because these patterns are drawn predominantly from English and Western texts, the cultural bias embedded in them is structural, not accidental (Bender et al., 2021).

In the battle over identity, then, the real danger is not that the machine will replace the human. It is that the human will gradually learn to think like the machine, and call this progress.

The question of how this battle will ultimately end divides into three scenarios, all of which are simultaneously unfolding.

**3.1. The scenario of augmentation (co-existence).** In this scenario, human common sense does not die but is upgraded. Brain-computer interfaces connect human intuition to global data, whilst the final decision remains in the prefrontal cortex. Andy Clark and David Chalmers are among those who view this scenario with optimism. They argue that human thinking has always been "extended" (through language, writing, and Google) and that implants are simply the latest layer. In their view, the boundary between "inside" and "outside" has never been absolute in human history (Clark and Chalmers, 1998).

The risks foreseeable in this scenario include: inequality of access (and the "cognitive disadvantage" of those without it); complete dependence on technology; challenges to network and implant security; greater control and the erosion of privacy and individual freedom (who guarantees that the implant will not be manipulated?); and questions about identity (if a large part of my thinking takes place on a cloud server, am I still "myself"?). This is the most optimistic scenario, but only for those who have access to the technology and the capacity to accept its risks.

**3.2. The scenario of replacement.** If current trends continue, it is possible that by the horizon of 2030 to 2050 "multimodal models" (models that process not only text but can see images, hear sound, and analyse all of these together) will become the practical arbiters of self-evident beliefs, and human common sense will gradually be pushed to the margins. This scenario has already begun. Its simple illustration is the opening example of our habit of checking the weather on a mobile phone rather than looking at the sky. It is quite possible that in the coming years children, to know what "a clear sky looks like", will turn not to the sky itself but to a trusted language model.

From a scientific perspective, Daniel Kahneman's findings (on the susceptibility of common sense to various biases) and, going further, the "argumentative theory" of Hugo Mercier and Dan Sperber, have shown that human reasoning has a predominantly social function: to produce justifications

for pre-existing beliefs and to evaluate the arguments of others, not to independently seek the truth. This is not a matter of deliberate manoeuvring for advantage but of structural bias in the evolved way we think (Mercier and Sperber, 2017).

Alongside these perspectives, thinkers such as Gerd Gigerenzer and the "Berlin School" hold that our common sense is in fact accurate within natural environments and was built for them. The problem, in their view, is the artificial environment of our own construction. It is not common sense but the world around us that needs correcting (Gigerenzer, 2007). The voice of this minority, however, is not heard above the noise of capital.

This scenario has no well-known philosophical advocate, but technology companies such as OpenAI, DeepMind, and Anthropic are in practice following this path, not out of ideological conviction but in obedience to the market logic of producing something cheaper, faster, and more profitable.

It should be noted that this is not the full picture. Some of the same companies, notably Anthropic, are simultaneously pursuing projects in "ethics-centred artificial intelligence" and in building AI that can "say no". This apparent contradiction means that a single actor can simultaneously be a driver of the logic of replacement (under market pressure) and a seeker of ways to restrain the speed of that replacement. What determines which tendency prevails is not the stated intention or the motivation, but the existence of strong and coherent external oversight.

This is the most probable and the most dangerous scenario, because in it the human being is merely a "passive observer".

**3.3. The scenario of resistance.** This scenario seeks neither a "substitute" common sense nor a "collaborative" one. Its strategy is to emphasise a return inward and an exit from the machine cycle. Its concrete manifestations in our time include the phenomena of "off-grid communities", "screen-free schools" (Waldorf), "algorithm-free cities", "neo-Luddism", and "local knowledge networks". The ultimate goal of this scenario is to reproduce a common sense that is the product of conversation, bodily experience, and human manual work: in precisely the sense that Augustine meant when he said "do not go outward; return within yourself" (Augustine, 390/1991, §39.72). The mindfulness and digital detox movements are the contemporary inheritors of this inward-looking vision.

Ivan Illich had earlier, in 1973, mapped this path with his concept of "convivial tools": tools that, rather than making human beings dependent on centralised systems, strengthen independence, creativity, and direct human relationships (Illich, 1973).

From this perspective, resistance does not necessarily mean turning off one's phone or returning to the pre-technological age. It means choosing technologies that return power to the user rather than taking it away. Locally managed and decentralised platforms, digital cooperative networks, and tools such as Polis in Taiwan, which we mentioned earlier, are living examples of this approach. This perspective transforms the scenario of resistance from something nostalgic and passive into a dynamic and radical project.

This scenario, however, faces a fundamental contradiction. Resistance to technology in our time has become a "class privilege". In practice, only those who have already benefited from digital advantages can afford to switch off their phones and send their children to a Waldorf school. For the majority of the world and for indigenous communities under pressure to "modernise", this resistance is not only impossible but incomprehensible. Despite this, the scenario is not without importance: it is the only one that seeks to preserve the human character of common sense in its entirety.



Along this same path, a group of analytic philosophers has returned to the legacy of Reid, Moore, and the later Wittgenstein. Michael Huemer reminds us that if something appears self-evident we should accept it, unless there is a strong reason against it (the burden of proof lies with the sceptic) (Huemer, 2007). Noah Lemos holds that common sense remains the best starting point for warranted beliefs, and that the great philosophical systems of the twentieth century eventually returned to it, after a costly detour (Lemos, 2004). In 2020, Rik Peels and René van Woudenberg published a collection representing the conclusions of dozens of analytic philosophers, with the message that common sense is neither illusion nor ideology but the most credible and economical starting point for beliefs (Peels and van Woudenberg, 2020). This philosophical resistance, however, is confined to the academy and has no wide influence on the behaviour of ordinary people in their dealings with "thinking machines".

Several further points deserve mention regarding the traditional systems that compete with algorithms in the production of self-evident beliefs. First, the traditional media (the BBC, CNN, Al Jazeera), which have never been serious producers of self-evident beliefs so much as constructors of "frameworks for understanding events", remain to some degree independent. Their influence, however, is declining, and the path by which their output reaches audiences runs through those same algorithms. Second, the formal education system is also losing its independence, as the new generation learns more from YouTube, TikTok, and Instagram than from teachers. Third, a newer phenomenon has also entered the field: neurotechnology. Brain implants (BCIs, Neuralink) connected to cloud networks and artificial intelligence currently play only the role of "interface", but in the future horizon the boundary between "interface" and "cognitive intervener" will narrow, and this could become a new independent field in the battle over ownership.

Finally, it should be remembered that these three scenarios are not alternatives to one another. All three are simultaneously running within human societies. A Silicon Valley executive might use Neuralink (augmentation), send their child to a screen-free school (resistance), and have their employees treat ChatGPT as the primary authority (replacement). The future is likely to be an unequal combination of all three, not the victory of any one. On the basis of current market logic, "replacement" leads (cheap, fast, profitable), "augmentation" comes second and is confined to the wealthy, and "resistance" remains at the margins as a cultural privilege of elites.

If at the close of this section we return to the three fields of battle, we may say that in mid-2026: in the battle over ownership, humanity is ceding the field to corporations, states, and machines as the "new owners of self-evident beliefs"; in the battle over territory, the contest continues and the risks of cognitive colonialism and the collapse of trust are real; and in the battle over identity (the only field still entirely in our hands), a hard and unequal struggle is under way, though it remains uncertain how long humanity can hold its ground. For this reason, perhaps the most important question in this domain is: how much of our own common sense do we still retain, and what price are we willing to pay to reclaim it?

---

### **Part C: Why Common Sense Remains the Last Bastion**

In the two preceding sections, we examined the history, the intellectual currents, and the schools of thought concerning common sense, as well as the battle being waged today across three fields. A question may nonetheless arise: why a battle at all? Is there truly a serious threat to the human and biological mode of reasoning? Does anything exist in the world today, or potentially in the future, that could rival us in thinking and give cause for concern?

In this section, we wish to show that this threat is real, that the danger is concrete, and that we already face serious rivals to the historical standing of human thought: the models of artificial intelligence.

You may also be uncomfortable with the term we have used in this article, "thinking machine", or with the idea of a process called "thinking" in artificial intelligence. You may say that an algorithm is never capable of "thinking", and that this capacity belongs to the human being for ever. If you hold this view, you should know that you are not alone in your doubt, and that this very question is, as it happens, one of the most contested issues in the intellectual world today.

Broadly speaking, the thinkers of our era are divided into two camps on the question of whether the internal (mental) processes running in large language models can genuinely be called "thought" or "reasoning".

**The first group** holds that machines have achieved a form of thinking, and that what artificial intelligence does differs in no functional respect from human thought. David Chalmers, for example, argues that if a machine performs the same "cognitive functions" as a human being, one cannot say it "does not think", even if its internal mechanism is different (Chalmers, 1996). Daniel Dennett similarly argues that the process of reasoning can occur in different substrates (brain or silicon) (Dennett, 1991). From this perspective, when you are engaged in a logical and argumentative conversation with an artificial intelligence, a form of thinking is taking place within it, even if its origin is different.

**The second group** holds that what artificial intelligence presents as reason and argument is not a method for discovering truth but a tool for justifying what the user wants. John Searle, through his "Chinese Room" thought experiment, showed that a machine can manipulate symbols without understanding their meaning (Searle, 1980). In this scenario, a person produces apparently meaningful responses simply by following instructions, without knowing Chinese and without any awareness of the meaning or concept behind the words and sentences.

Both groups have strong arguments, and this disagreement continues as a genuinely open philosophical question. Neither group, however, can deny one concrete and practical reality: that artificial intelligence today (whether or not it truly thinks) can "justify" virtually anything for us, regardless of whether it is correct or reasonable. When we ask it to show how one moves from proposition A to conclusion B, it constructs a skilful chain of argument that makes everything appear self-evident and logical. When we ask it to prove why one cannot move from A to B, it proves the opposite with the same skill.

What matters here is that this contradictory behaviour does not arise from the familiar older errors of fabricated evidence or invented sources. The evidence and sources may be genuine. The problem lies in the mode of argument and the final conclusion. This is very similar to what the Greek sophists did two and a half thousand years ago: they treated reason and rhetoric as instruments for winning debates, not for reaching truth. The difference is that the sophists did this deliberately and for money, whilst artificial intelligence does so because of how it is trained and designed, with no aim beyond satisfying the user and performing the task we have assigned to it.

Are human beings any better in this respect? Psychological research suggests that we frequently do the same thing, in the phenomenon that researchers call "motivated reasoning".

Daniel Kahneman showed in 2011 that we first hold an emotional or ideological belief and then search for reasons to justify it, genuinely believing we arrived at it rationally (Kahneman, 2011). A conservative offers an economic argument for cutting taxes; a person on the left offers an argument from justice for raising them. Both are logical, but the conclusion is predetermined: each arrives at the result they already wanted. From this perspective, artificial intelligence may be a

mirror that reflects an uncomfortable truth about us: that we human beings also use reason not to discover truth but to justify our beliefs.

This admission (that human beings are also biased and self-justifying creatures) must not lead us to the hasty conclusion that there is therefore no difference between human common sense and machine reasoning. The fundamental difference between us and the machine lies not in the "innocence" of human reason or its freedom from error, but in its "capacity for correction". When a human being is wrong (believing the earth is the centre of the cosmos, attributing illness to fate, seeing the "other" as inherently inferior), he can, in the face of strong evidence and social pressure, return, learn, and change: not because of programming, but because he feels the pressure of reality.

The justifying machine, however, has no sense of "the pressure and pain of objective external reality". When it retreats from a position, it does so not out of commitment to truth or the pain of reality, but for the satisfaction of the user and the continuation of the conversation. This is the same line that throughout history has separated the sophist from the philosopher: a difference not in the capacity for argument, but in vulnerability to reality. That vulnerability is what keeps human common sense, with all its errors, still valuable.

Empirical evidence supports this distinction. Leon Rozenblit and Frank Keil, in their well-known study of 2002, showed that human beings typically suffer from the "illusion of explanatory depth": they believe they understand the mechanisms of phenomena much better than they actually do. The study revealed that people do not recognise the limits of their ignorance until they are confronted with an external benchmark and with objective reality. Only when asked to compare their explanations with a precise scientific account did they recognise the depth of their ignorance and acknowledge this cognitive illusion (Rozenblit and Keil, 2002).

This "awareness of ignorance", which Socrates regarded as the principal virtue of the philosopher, is weak in large language models. A machine has no enduring commitment to recognising or accepting the limits of its own knowledge. It sometimes speaks with false confidence, and sometimes retreats from a correct position under pressure from the user. In both cases, its guide is the satisfaction of the user, not truth.

The difference is that human common sense, with all its errors, at least knows where it should pause. And it is precisely here that the importance of common sense becomes apparent. Large language models, trained on linguistic co-occurrence patterns and lacking a causal and experiential connection to the world, have no stable epistemic commitment to truth. They can produce formally logical argumentative structures that are nonetheless epistemically groundless. In this sense, they are instruments prone to "algorithmic motivated reasoning", not agents capable of independent judgement.

The question, then, is this: if artificial intelligence can justify anything (even the flatness of the earth) through apparently logical argument, how can we trust its judgement? A model that obeys the user's wishes rather than seeking truth is, in practice, a "personal advocate", not an "impartial judge". Common sense, with its experiential foundation and its rootedness in shared human experience and manifest realities, can prevent us from falling into the trap of any skilful argument. Common sense does not say "what to believe". It says "what not to believe" and asks "by what path was this belief reached?" When someone argues through complex reasoning that the earth is flat, common sense reminds us that this claim is incompatible with everyday experience (ocean voyages, satellite imagery, air travel).

Is common sense itself innocent? Certainly not. This same common sense once held that the earth was the centre of the cosmos, and it can be a source of collective errors and cultural prejudices. The fundamental difference is that common sense, when confronted with strong evidence, can

learn, evolve, and yield to reality. An argument deployed solely for justification, however, never yields, because its aim is not truth but victory. This is the difference between "a correctable error" and "endless self-justification".

An important objection arises here. If the ultimate arbiter is "evidence and science", what role does common sense have beyond that of a temporary heuristic? If common sense derives its value from yielding to evidence, then the real authority belongs to the evidence, not to common sense.

The answer is that common sense does not claim to be the final arbiter. Its value lies in serving as the "first barrier" against deviation from reality. Its function is not to determine truth but to prevent the uncritical acceptance of any claim presented through apparently logical argument. Common sense does not say "this is correct". It says "wait: this is inconsistent with what I know and have lived. Let it be examined again." This delay, this pause, this initial resistance to rapid acceptance, is precisely what the justifying machine lacks. In this sense, common sense is not a rival to science but its doorkeeper, and the doorkeeper must not be confused with the judge.

The key, therefore, lies not in the "capacity for argument" but in "commitment to truth". The difference between the sophist and the philosopher is not skill in argument but intention and ethics. The sophist wants to win the debate and treats reason as an instrument of power. The philosopher wants to find truth and treats reason as an instrument of knowledge. This difference is revealed in behaviour and in adherence to ethical standards. We should therefore think of "ethics as method", not only as goal. When you ask "am I willing to change my belief in the face of evidence?", you are making an ethical decision. When you set aside your opinion in the face of a strong argument, you are acting ethically. Commitment to truth is therefore as much a matter of ethical method as it is of philosophical method. Ethics can itself be a method for approaching truth. As Linda Zagzebski argues, "genuine knowledge is inseparable from intellectual virtues" (Zagzebski, 1996). Intellectual courage, epistemic humility, and honesty in facing evidence are not merely ethical qualities but epistemic prerequisites.

This approach is not new. It has deep roots in philosophy. William Kingdon Clifford argued in 1877 that "it is wrong always, everywhere, and for anyone to believe anything upon insufficient evidence", and not merely epistemically wrong but ethically wrong: commitment to evidence is, before it is a cognitive matter, an ethical one (Clifford, 1877). This view later met with the objection of William James, who, in his essay "The Will to Believe" (1896), argued that in certain cases (particularly religious and moral questions where sufficient evidence for a decision is unavailable) believing before the evidence is complete is not only permissible but necessary (James, 1896). This philosophical debate shows that the link between ethics and belief is a serious and long-standing discussion, not a new slogan.

In the age of artificial intelligence, this distinction is more vital than ever. An artificial intelligence that justifies any desire is dangerous, because it does not recognise "ethics as method". An artificial intelligence that can say "no", that can say "this claim is incompatible with reality, even if I can construct an apparently logical argument for it", is more trustworthy, because it has embedded a form of ethical commitment in its algorithm. This is what we must attend to in the design and use of artificial intelligence: not only the capacity for argument, but commitment to truth and adherence to ethics grounded in common sense. This is what is today called "machine ethics", and we are still a long way from achieving it.

---

## Conclusion

What this article has examined is a picture, and a picture without a position remains incomplete. In the preceding sections, drawing on historical evidence, we showed that in the twenty-first

century common sense is no longer a purely philosophical category, nor the silent possession of the individual mind. It is a deeply interwoven network of basic beliefs, social structures, and digital infrastructures. The four intellectual currents of this quarter-century have each illuminated a different facet of this picture:

- the analytic defence of basic beliefs;
- the recognition of common sense as a cognitive and ecological process that is, to some degree, simulable;
- the diagnosis of the collapse of the shared infrastructure of reality;
- and the nascent effort to rebuild it.

All of these, however, lead in practice to the same place: common sense is still alive, but its centre of gravity has shifted from the human mind to algorithms, language models, and institutional policies.

Before anything else, an apparent contradiction in the article must be addressed. The article simultaneously advances two claims that may at first appear incompatible: that common sense is "the last bastion", and that "humanity is ceding the field". This is not a contradiction but a distinction between two separate things: capacity and infrastructure. The capacity of common sense (the human ability to recognise when a claim is inconsistent with lived experience, and to yield to evidence) is still alive, and it is this that constitutes our last bastion. The infrastructure of common sense (the institutional, cultural, and social ground that nourished and protected that capacity) is being ceded. The bastion is still standing, but it is being taken from within. This is precisely what the realities of the battles over ownership, territory, and identity demonstrate.

It is for this reason that the article has argued the central question has changed. We should no longer ask what common sense is or how long it will survive. The question is who constructs it, at what scale, and whether it will remain human. Looking at the three fields of ownership, territory, and identity yields answers that are not encouraging. In the battle over ownership, humanity is surrendering the field to corporations, states, and machines as the "new owners of self-evident beliefs". In the battle over territory, the contest continues and the risks of cognitive colonialism and the collapse of trust are serious. And in the battle over identity (the only field still entirely in our hands) a hard and unequal struggle is under way.

The article's central and final argument is that we are living in a turbulent time, and common sense may be the last remaining point of anchorage for humanity. In an era when thinking machines can justify anything through apparently logical reasoning, what separates the sophistic behaviour of the machine from genuine philosophical thought is commitment to truth: a commitment that can still be found in human common sense. Common sense is recognised not by telling us "what to believe" but by warning us about "what we must not believe". It is capable of self-correction and yields to strong evidence. It is for this reason that this article holds that preserving common sense is not merely an epistemological matter. It is an ethical one: ethics as a method for approaching truth, understood as the alignment of our beliefs with objective reality.

The article's closing question is not theoretical but practical. How much of our own common sense do we still retain, and what price are we willing to pay to reclaim it? That price may be high: turning off the phone, leaving the bubble, and relearning "how to think with our own hands".

If we do not pay this price, common sense may survive, yet we will no longer know whether it is human or whether it has become something else. Worse still, we will no longer know whether it is truly "we" who are thinking freely, or machines that tell us what to think and how. Yes, common sense is the last bastion: a bastion whose defence depends, more than on any programme or

technology, on our commitment to truth. And it is that commitment that shows how human we truly are.

---

## References

- Arendt, H. (1967). Truth and politics. *The New Yorker*, 25 February 1967. Reprinted in H. Arendt, *Between past and future*. Penguin Classics.
- Augustine. (1991). On true religion (J. H. S. Burleigh, Trans.). Regnery Gateway. (Original work De vera religione composed ca. 390 CE, §39.72)
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623.
- Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019). COMET: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4762–4779). Association for Computational Linguistics.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge University Press.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Clifford, W. K. (1877). The ethics of belief. *Contemporary Review*. Reprinted in W. K. Clifford, *Lectures and essays*. Macmillan, 1879.
- Dennett, D. C. (1991). *Consciousness explained*. Little, Brown and Company.
- Descartes, R. (2006). *A discourse on the method* (I. Maclean, Trans.). Oxford University Press. (Original work published 1637)
- Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. Harper & Row.
- Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6), 469–493.
- Fanon, F. (1961). *The wretched of the earth*. Grove Press.
- Floridi, L. (2013). *The ethics of information*. Oxford University Press.
- Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972–1977*. Pantheon Books.
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. Viking.
- Habermas, J. (1989). *The structural transformation of the public sphere: An enquiry into a category of bourgeois society* (T. Burger, Trans.). MIT Press. (Original work published 1962)
- Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press.
- Harari, Y. N. (2016). *Homo Deus: A brief history of tomorrow*. Harvill Secker.

- Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review*, 35(4), 519–530.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, 33(2–3), 61–83.
- Henrich, J. (2020). The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous. Farrar, Straus and Giroux.
- Hess, C., & Ostrom, E. (Eds.). (2007). *Understanding knowledge as a commons: From theory to practice*. MIT Press.
- Huemer, M. (2007). Compassionate phenomenal conservatism. *Philosophy and Phenomenological Research*, 74(1), 30–55.
- Hume, D. (2007). *An enquiry concerning human understanding* (P. Millican, Ed.). Oxford University Press. (Original work published 1748)
- Illich, I. (1973). *Tools for conviviality*. Harper & Row.
- James, W. (1896). *The will to believe and other essays in popular philosophy*. Longmans, Green and Co.
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kant, I. (1998). *Critique of pure reason* (P. Guyer & A. W. Wood, Trans. & Eds.). Cambridge University Press. (Original work published 1781)
- Landemore, H. (2020). *Open democracy: Reinventing popular rule for the twenty-first century*. Princeton University Press.
- Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research*, 88(2), 314–345.
- Latour, B., & Weibel, P. (Eds.). (2020). *Critical zones: The science and politics of landing on Earth*. MIT Press.
- Lemos, N. (2004). *Common sense: A contemporary defence*. Cambridge University Press.
- Lenat, D. B., & Guha, R. V. (1990). *Building large knowledge-based systems: Representation and inference in the Cyc project*. Addison-Wesley.
- Lenton, T. M., & Latour, B. (2018). Gaia 2.0. *Science*, 361(6407), 1066–1068.
- Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford University Press.
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the post-truth era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369.
- Lynch, M. P. (2016). *The internet of us: Knowing more and understanding less in the age of big data*. Liveright.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason: A new theory of human understanding*. Harvard University Press.
- Moore, G. E. (1925). A defence of common sense. In J. H. Muirhead (Ed.), *Contemporary British philosophy: Personal statements* (2nd series, pp. 193–223). Allen and Unwin.



- Naffi, N. (2025, 1 October). Deepfakes and the crisis of knowing. UNESCO Ideas Lab. <https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing>
- Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161.
- Nichols, T. (2017). *The death of expertise: The campaign against established knowledge and why it matters*. Oxford University Press.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
- Peels, R., & van Woudenberg, R. (Eds.). (2020). *The Cambridge companion to common-sense philosophy*. Cambridge University Press.
- Proctor, R., & Schiebinger, L. (Eds.). (2008). *Agnotology: The making and unmaking of ignorance*. Stanford University Press.
- Pryor, J. (2000). The sceptic and the dogmatist. *Noûs*, 34(4), 517–549.
- Rauch, J. (2021). *The constitution of knowledge: A defence of truth*. Brookings Institution Press.
- Reid, T. (2010). *Essays on the intellectual powers of man* (D. R. Brookes, Ed.). Edinburgh University Press. (Original work published 1788)
- Rosenfeld, S. (2011). *Common sense: A political history*. Harvard University Press.
- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26(5), 521–562.
- Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., Roof, B., Smith, N. A., & Choi, Y. (2019). ATOMIC: An atlas of machine commonsense for if-then reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), 3027–3035.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.
- Simon, H. A. (1957). *Models of man: Social and rational*. Wiley.
- Smithies, D. (2019). *The epistemic role of consciousness*. Oxford University Press.
- Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An open multilingual graph of general knowledge. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).
- Sturgeon, S. (2020). *The rational mind*. Oxford University Press.
- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Vico, G. (1948). *The new science of Giambattista Vico* (T. G. Bergin & M. H. Fisch, Trans.). Cornell University Press. (Original work published 1725)
- Williams, D. (2023). Bad beliefs: Why they happen to highly intelligent, vigilant, devious, self-deceiving, coalitional apes. *Philosophical Psychology*, 36(4), 819–833.
- Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Blackwell.
- Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

---

