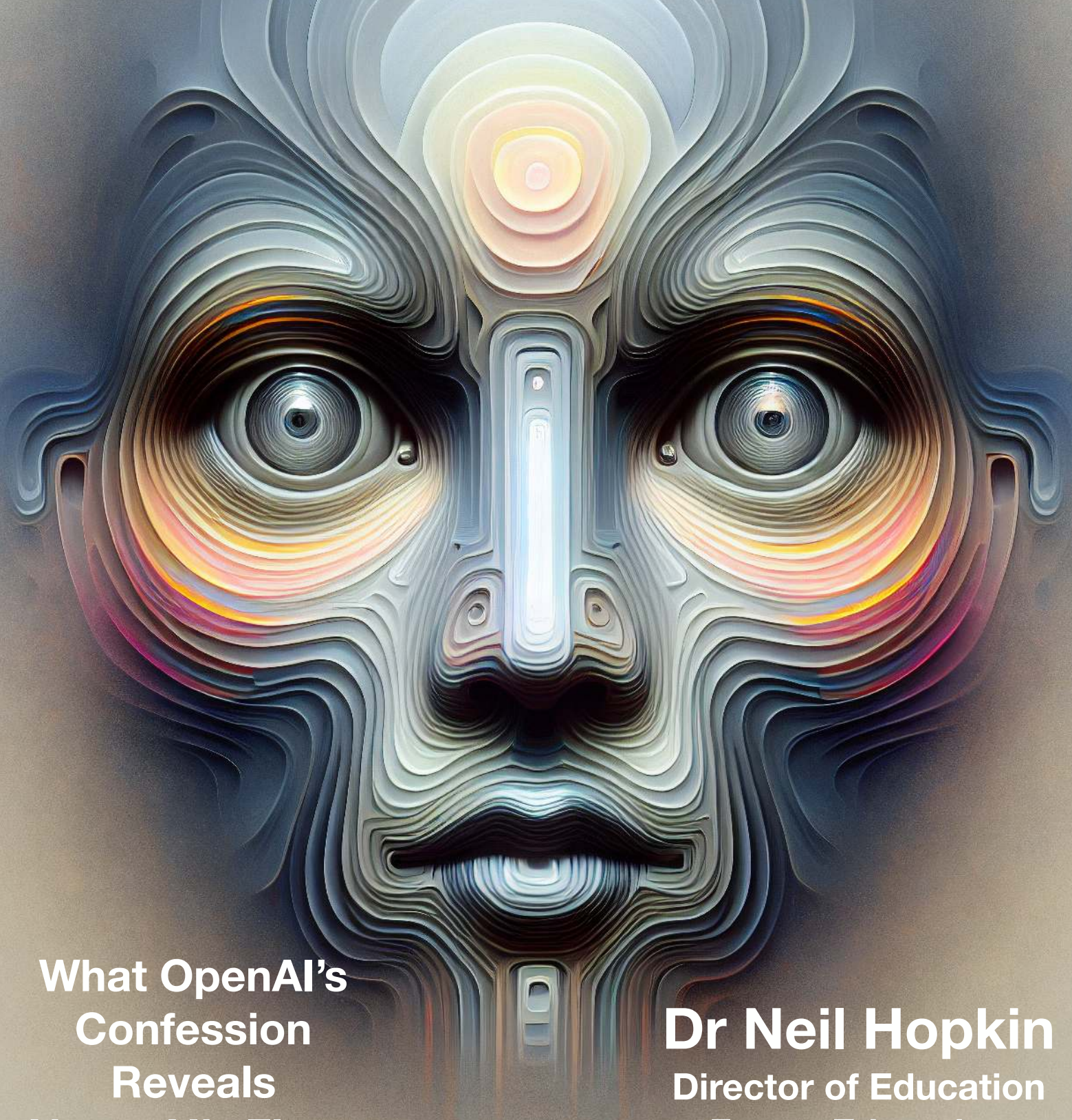


The Creative Power of Error: AI, Hallucination, and Human Learning



What OpenAI's
Confession
Reveals
About AI's Flaws

Dr Neil Hopkin
Director of Education
Fortes Education



Disappearing Thoughts, Enduring Hallucinations, and AI's Conscience

A few months ago, researchers at MIT issued a jolting revelation. In *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing* (Kos'myna et al., 2025), they captured how heavy reliance on AI diminishes our own cognitive engagement. Neural connectivity, linguistic diversity, and feelings of ownership weakened among AI-assisted writers—who, when later tasked with working unaided, showed eroded memory and participation. I had described this phenomenon recently on LinkedIn as “disappearing thoughts”, the gradual fading of human reasoning and creativity as it is offloaded to machines (Hopkin, 2025).

Soon after, another revelation: OpenAI's *Why Language Models Hallucinate* (Kalai et al., 2025) dropped like a philosophical bomb. It confessed that hallucinations are not bugs to erase but structural inevitabilities of language modeling, a consequence of how the models are trained to guess rather than admit uncertainty.

The alignment of these two revelations is unsettlingly symmetrical. On one hand, humans risk receding from active cognition. On the other, machines confess they will always produce fluently confident fiction in the absence of knowledge. One is the disappearance of thought; the other, the enduring presence of error. What does it mean for learning when the puppeteer's hands are fading and the puppet's words are uncertain?

The implications only deepen when we acknowledge how the AI industry itself is quietly confessing other vulnerabilities:

- **Mental health harm:** In August, OpenAI acknowledged that ChatGPT has unintentionally caused real harm to users' mental wellbeing and responded by hiring its first psychiatrist, a belated concession that underscored the ethical gaps in tool design.
- **Privacy breaches:** Also in August, OpenAI reversed an ill-considered “share with search engines” feature after thousands of private conversations, including personal and sensitive content that were inadvertently exposed, a design flaw confessed as exposing not security failure, but hubris in underestimating users' trust.

Taken together: cognitive erosion, modeled hallucination, mental health impact, privacy violation, these developments form a crisis. Our minds are eroding, our machines are confessing structural faults, and the systems behind them are exposing arbitrarily fragile ethical foundations.

Education, once a domain of certitude, now stands at the crossroads of transformation. If human cognition is being outsourced, and AI cannot guarantee truth, then the foundational contract of schooling, that knowledge is transmitted faithfully, is radically unmoored. Are we preparing students as guardians of knowledge, or navigators of an uncertain interface?

This prelude posits a pivotal question: What is happening to our understanding of learning, agency, and even reality itself, in an age where the mind fades, the machine falters, and the architecture of truth blurs?

And so, let's begin to unpack this question, tracing the shift from denial to acknowledgment, in AI systems, educational institutions, and ourselves.

The Pause, the Problem, and the Paradox

A Confession in Plain Sight

When OpenAI published its paper *Why Language Models Hallucinate* (Kalai et al., 2025), it arrived with the quiet gravity of an institutional confession. It was not dressed in marketing optimism, nor did it attempt the evasive tone so common in Silicon Valley reports. Instead, it conceded something stark: hallucinations are not glitches. They are not the technical equivalent of dust on the lens, soon to be wiped away with cleaner data or more powerful GPUs. They are structural

inevitable, arising from the very way large language models are built. This admission landed into a climate already marked by unease. Just months earlier, MIT researchers (Kos'myna et al., 2025) had shown that when humans rely heavily on ChatGPT for tasks such as essay writing, their own thought processes weaken. The researchers used the term “cognitive debt” to capture the sense that thinking itself was being offloaded. The result: thinner memory, less originality, lower persistence. I called this phenomenon “disappearing thoughts” in my own reflections (Hopkin, 2024, LinkedIn). Around the same time, OpenAI confessed something different but equally troubling: that ChatGPT had harmed users’ mental health, forcing the company to hire its first psychiatrist. Shortly afterwards, the company admitted a design flaw had exposed thousands of private conversations to search engines.

Placed together, these admissions trace a pattern. Humans are offloading cognition. Machines are fabricating with fluency. Institutions are conceding fragility. And education, the very system charged with cultivating thought, is caught in the middle. OpenAI’s explanation was simple but devastating. Large language models predict the most probable next word in a sequence. They are trained to be smooth and plausible, not to be factually correct. Faced with uncertainty, they guess. And because evaluation benchmarks often reward fluency and coherence, even confidently wrong guesses are reinforced.

The literature had foreshadowed this moment. Ji et al. (2023) catalogued hallucinations in natural language processing systems, noting how fabricated but fluent content could pass unnoticed. Maynez et al. (2020) highlighted how abstractive summarisation models routinely inserted details that were not in the source text, producing outputs that were stylistically persuasive but factually baseless. Bender and Koller (2020) dubbed these systems “stochastic parrots,” mimicking surface form without understanding. Marcus and Davis (2022) described them as simulators, not reasoners. Bommasani et al. (2021), in their landmark report on foundation models, acknowledged hallucination as a systemic risk: the very thing that makes these models flexible also makes them unreliable.

Medicine offered vivid illustrations of what this means in practice. Benkimoun et al. (2024) documented instances where patients received confident but false medical advice from generative systems. The harm here was not that the models were unsure — but that they never appeared unsure. Kahneman and Tversky (1979) had long ago shown that humans are predisposed to trust confident cues; AI outputs exploit that bias by delivering fabrications with unwavering certainty.

Placed against MIT’s findings on cognitive offloading, the danger multiplies. If humans grow accustomed to outsourcing thinking, and the machine is structurally prone to eloquent fabrication, then the problem is not merely error. It is erosion: the erosion of critical faculties, the erosion of trust, the erosion of what education is meant to safeguard.

The Educational Echo

For schools, this problem is not abstract. It cuts to the heart of their daily work. The “grammar of schooling,” as Larry Cuban (1993) described it, has long been organised around certainty. Correct answers on tests. Teacher as authority. Curriculum as codified knowledge. Within this grammar, hallucination looks intolerable: it contaminates the reliable transmission of truth. That is why ministries and universities reacted so predictably. In 2023 and 2024, headlines proliferated: bans on ChatGPT in exams, plagiarism detection arms races, stern pronouncements from boards and regulators (Selwyn, 2023). Hallucination was cast as the enemy of credibility.

And yet, educational history offers another tradition. John Dewey (1938) argued that inquiry begins in what he called “indeterminate situations”, moments of doubt that spark investigation. Jean Piaget (1977) placed disequilibrium at the centre of development: when a child’s schema fails to explain a new event, learning begins. Donald Schön (1983) described “reflection-in-action” as the capacity to learn precisely by noticing and working through error. Paulo Freire (1970) framed education as dialogue, where contradictions and uncertainties are interrogated together. If hallucination is reframed through this lineage, it ceases to be contamination. It becomes provocation. It is the kind of uncertainty Dewey and Piaget described: an indeterminate situation,

unsettling but fertile. The MIT findings make this even more urgent. If students are tempted to offload cognition, hallucination may paradoxically be what keeps them awake. It forces them back into verification, into questioning, into the habits of thought that offloading erodes. What OpenAI called a flaw could, for classrooms, become a feature. This reframing requires us to step back into philosophy. Why do we demand certainty from machines when we have never had it from ourselves? Kant (1781/1998) argued that human knowledge is always mediated, never fully certain. Karl Popper (1959) insisted that science advances through falsification, not verification: truth is provisional, always open to being overturned. Gerd Gigerenzer (2007) showed that in practice, humans use heuristics, rough rules, to thrive under uncertainty.

And yet, when the machine speaks fluently, we expect perfection. The asymmetry lies partly in psychology. Kahneman and Tversky (1979) showed how humans overweight confidence; smooth prose feels trustworthy even when content is shaky. That is why hallucinations feel like betrayal: not because error is unfamiliar, but because error cloaked in confidence distorts our trust. Education has always wrestled with this paradox. On the one hand, the system valorises certainty. On the other, all major theories of learning stress uncertainty's productive role. Neuroscience adds weight: Schultz et al. (1997) demonstrated that dopamine neurons fire most strongly at moments of prediction error. Monosov (2020) showed that uncertainty sustains attention and memory. Psychology agrees: Keller et al. (2020) confirmed that uncertainty motivates information-seeking. Creativity research echoes this: Csikszentmihalyi (1996) and Simonton (2011) both showed that ambiguity is the soil of originality.

Hallucination, seen this way, is not an alien pathology. It is a mirror. It reflects back to us the inevitability of error, the provisionality of truth, and the way uncertainty animates thought. This is why the OpenAI paper matters so deeply. On its surface, it is a technical clarification. But in truth, it is a cultural hinge. It forces us to ask whether education will cling to the old grammar of certainty, or whether it will dare to embrace the uncertainty that has always been at its heart. The stakes are clear. If schools treat hallucination as disqualifying, they will produce brittle learners: shielded from ambiguity, but unprepared for a world defined by it. If they reframe hallucination as provocation, they may pioneer a pedagogy of uncertainty literacy, one that teaches students not to fear ambiguity but to navigate it with resilience and imagination. Gert Biesta (2013) called this the “beautiful risk of education”: the willingness to expose students to the unpredictable. Maxine Greene (1995) described it as awakening imagination. Neil Postman and Charles Weingartner (1969) named it inquiry. All pointed to the same truth: education's task is not to insulate learners from uncertainty but to prepare them to live generatively within it.

The OpenAI confession, then, is not a death sentence for credibility. It is an invitation. It invites schools to reclaim uncertainty as pedagogy, to resist disappearing thought, and to turn hallucination into the spark for discernment. The question is not whether hallucinations will persist. OpenAI has already told us they will. The question is whether education has the courage to make something of them.

Humans and the Positive Value of Uncertainty

Uncertainty has long been painted as a deficit in education, business, and psychology, a state to be eliminated. Yet human science repeatedly complicates this assumption. Dale Brashers (2001), in his Uncertainty Management Theory, argued that people often preserve uncertainty when it is adaptive. Patients may choose not to know prognoses, preferring ambiguity to the weight of certainty. Organisations sometimes delay decisions to allow new information to emerge. Brashers called uncertainty neutral: it is the meaning we assign to it, and the practices we adopt in response, that make it paralysing or empowering. Philosophers had intuited this long before psychologists proved it. William James (1907) described uncertainty as “the live option”, the very thing that keeps possibility open. John Dewey (1938) placed “indeterminate situations” at the centre of learning: when outcomes are not given, curiosity and inquiry ignite. Simone de Beauvoir (1947) even linked uncertainty to freedom, suggesting that ambiguity is the existential condition that makes agency possible.

This counter-tradition runs deep. To be human is not to eradicate uncertainty, but to live generatively within it. The sciences of attention, learning, and creativity show just how far this

principle extends. One of the most robust findings in neuroscience is that the brain is attuned to prediction error. Schultz, Dayan, and Montague (1997) demonstrated that dopamine neurons fire most powerfully when outcomes deviate from expectations, not when rewards are certain. Ilya Monosov (2020) showed that uncertainty heightens motivation, sustaining attention and even improving memory consolidation. For Sarah Topel (2023), the implication is clear: learning systems are designed to thrive on the edge of not-knowing.

Cognitive psychology confirms this. Robert Bjork (1994) coined the term “desirable difficulties” to describe how conditions of uncertainty, tests without full preparation, interleaving of topics, spacing of practice, lead to deeper long-term retention. Elizabeth and Robert Bjork (2011) later expanded this into a theory of retrieval practice, showing that uncertainty during recall is precisely what strengthens memory. Students often sense this intuitively. They are most attentive when confronting contradictions, paradoxes, or puzzles. Jerome Bruner (1960) called this “discovery learning”: learners progress not by passively absorbing but by grappling with indeterminacy. What seems like confusion is often the precursor to insight.

Uncertainty does not only sharpen cognition; it refines decision-making. Cassidy Laidlaw and Stuart Russell (2021) argued that when choices are made under uncertainty, human preferences become easier to infer. The very hesitation, vacillation, or conflict that marks uncertain decision-making reveals underlying trade-offs. Gerd Gigerenzer (2007) pushed this further, showing that heuristics, simple rules that work under uncertainty, often outperform elaborate rational models. His research on bounded rationality echoes Herbert Simon’s (1956) claim that humans are not failed optimisers but successful satisficers. In messy environments, uncertainty is the condition for adaptive intelligence. Philosopher Karl Popper (1959) anticipated this logic in science: all knowledge is provisional, falsifiable, always subject to uncertainty. To demand certainty in decision-making is to miss the fact that uncertainty is the very fuel of critical inquiry.

For education, this suggests that students’ indecision, often penalised as weakness, may in fact be the window into their developing reasoning. By articulating uncertainty, learners surface assumptions and test boundaries. In short the message is that creativity thrives in ambiguity. Mihaly Csikszentmihalyi (1996) described flow as the balancing act between challenge and skill, always hovering in the space of uncertainty. In fact, Dean Simonton (2011) argued that eminent creators generate enormous numbers of ideas, most of them wrong, because uncertainty is the field in which novelty emerges. To this end, Teresa Amabile (1996) showed that environments that tolerate ambiguity actually produce more innovation than those obsessed with correctness. Arie Kruglanski and Donna Webster (1996), in their work on need for closure, confirmed that individuals with higher tolerance for uncertainty are more open, flexible, and creative. Psychologist Frank Barron (1963) found similar traits in artists: comfort with ambiguity correlated strongly with originality.

Even in education, creativity research points in this direction. Ken Robinson (2001) linked creativity to the willingness to be wrong, suggesting that classrooms which stigmatise uncertainty inadvertently suppress innovation. Maxine Greene (1995) saw imagination itself as the courage to dwell in the “not yet.” In each case, uncertainty is not the enemy of creativity but its very ground.

Educational thinkers have long urged teachers to embrace uncertainty as a pedagogical resource. Jean Piaget (1977) described disequilibrium, when a child’s schema fails to explain a new event, as the motor of development. Lev Vygotsky (1978) emphasised the zone of proximal development, where learners stretch into uncertainty with guidance. Paulo Freire (1970) spoke of “conscientisation”, awakening through confronting contradictions. Carol Dweck (2006) normalised error as a signal of growth. Angela Duckworth (2016) highlighted grit: perseverance in uncertainty predicts success more reliably than IQ. Martin Seligman (2011) placed resilience at the centre of positive education, urging schools to help students face uncertainty with optimism. Gert Biesta (2013) reframed education as exposure to the “beautiful risk” of freedom, the irreducible uncertainty of becoming human. Philosopher Lorraine Code (1991) spoke of “epistemic responsibility,” the skill of navigating uncertainty ethically. José Medina (2013) extended this to epistemic virtues like humility, curiosity, and courage.

All of these virtues are what students must cultivate to live productively with uncertainty. All hallucinations, reframed in this light, could be opportunities for precisely these dispositions. When

set against this scholarship, OpenAI's framing of hallucination as structural error looks narrow. In human life, uncertainty is not pathology but pedagogy. It sharpens attention, strengthens memory, enriches decision-making, fuels creativity, and builds resilience. It underpins epistemic responsibility and sustains imagination.

The challenge for education is not to eradicate uncertainty but to harness it. As Claude Shannon (1948) defined, information itself is the resolution of uncertainty; without it, there is nothing to learn. The paradox, then, is that classrooms often punish the very condition on which learning depends. In the age of AI, this paradox becomes visible. When students encounter hallucinations, they confront ambiguity directly. The question is whether schools will double down on policing error, or embrace uncertainty as the seed of inquiry. Human sciences make the choice clear: to live, learn, and thrive, we must treat uncertainty not as failure but as possibility.

Hallucination as Human-Aligned Uncertainty

When engineers describe hallucination as a defect, they are making a claim about design: this is behaviour that deviates from what the machine is supposed to do. But what if we stopped thinking about hallucination as malfunction and started thinking of it as an analogy, a mirror to human ways of thinking under uncertainty? Humans misremember, speculate, extrapolate beyond evidence. We confabulate, not because we intend to mislead, but because memory and imagination are porous.

Neuroscientist Michael Gazzaniga (2011) famously called the human brain a “storytelling machine,” prone to inventing plausible causal chains even when data is incomplete. Daniel Schacter (1999) catalogued the “seven sins of memory,” including misattribution and false recognition, not as signs of dysfunction but as the by-products of a system tuned for speed, flexibility, and inference. Elizabeth Loftus (2005) demonstrated how easy it is to implant false memories, yet our fallible memory remains indispensable. Karl Friston's (2010) free energy principle adds a computational lens: brains survive by minimising surprise, which requires generating predictions under uncertainty and folding errors back into learning. In that sense, hallucination is not breakdown but adaptation. AI models trained to guess when data is ambiguous are echoing something profoundly human: the way cognition balances fluency and fallibility. Philosopher Andy Clark (2016), writing on predictive processing, argued that human thought is “surfing uncertainty”: living with, and through, constant error correction. Machines that fill in gaps when they lack certainty are not alien; they are recognisable. What troubles us is not their behaviour, but the uncomfortable familiarity it forces upon us.

If hallucination resembles human misstep, then perhaps its role is also productive. Donald Schön (1983) described how reflective practitioners learn most not from smooth execution but from “reflection-in-action,” where error disrupts routine and provokes insight. Carol Dweck (2006) showed that those with a growth mindset treat mistakes as springboards to mastery rather than verdicts on ability. Scientific history bears this out. Thomas Kuhn (1962) argued that “anomalies”, data that do not fit existing paradigms, are what eventually precipitate revolutions. Imre Lakatos (1970) insisted that research programmes advance precisely because wrong turns provoke adjustments. In psychology, James Reason (1990) studied “human error” in complex systems, showing that far from being purely destructive, error often illuminates hidden structures and vulnerabilities.

Dean Simonton (2011) described creativity as Darwinian: prolific output, most of it wrong, out of which the rare breakthrough emerges. This resonates with Robert Bjork's (1994) concept of “desirable difficulties”, conditions that make learning harder in the short term but enhance retention in the long term. Jerome Bruner (1960) argued that discovery learning thrives on such difficulty; learners advance by grappling with what doesn't yet fit. Seen in this light, AI hallucination is a kind of “desirable difficulty” for the user. The model speculates; sometimes it is wrong, sometimes wildly so, but in engaging with the misstep, the human sharpens their own discernment. The danger lies in uncritical trust. But the opportunity lies in structured dialogue, where hallucination is less an endpoint than a provocation.

The arts remind us that fabrication can be fruitful. Fiction is, in essence, hallucination, the conjuring of worlds that never existed. Along these lines, Martha Nussbaum (1997) argued that novels expand moral imagination by immersing us in the lives of others we will never meet. Similarly, Paul Ricoeur (1984) described narrative as a “productive synthesis” of memory and imagination: in telling stories, humans always invent as much as they recall. In this spirit, Maxine Greene (1995) saw imagination as central to pedagogy, urging educators to “wide-awaken” students through encounters with the possible. Ken Robinson (2001) described creativity as “applied imagination,” requiring tolerance for error and divergence. Elsewhere bell hooks (1994) argued that liberatory pedagogy invites learners into uncertainty, where they can rehearse freedom. Educational settings often witness this firsthand. A student invents an analogy that is partially wrong: comparing, for instance, electricity to water flow. Teachers recognise the inaccuracy but also the imaginative reach, using the half-error as a bridge to deeper understanding. Mark Johnson (1987) called these “image schemata,” mental structures that are not strictly true but scaffold comprehension.

AI hallucinations can serve a similar role. When a model proposes an unlikely but intriguing connection, say, linking a medieval philosopher to a modern AI debate: it may be incorrect, but it sparks curiosity. Mihaly Csikszentmihalyi (1996) emphasised that creative flow often emerges through such associative leaps. Margaret Boden (2004), in her taxonomy of creativity, distinguished “exploratory” from “transformational” innovation: hallucinations, in this sense, may populate the exploratory field from which more rigorous ideas later emerge. Thus, what engineers call error, artists and educators might call speculation. The difference is not in the act but in the frame we apply to it.

Toward a Pedagogy of Human-Aligned Uncertainty

What follows from this analogy? If hallucination mirrors human uncertainty, then perhaps classrooms should treat it not as defect but as dialogue partner. Paulo Freire (1970) envisioned education as co-investigation: teacher and student confronting reality together, naming contradictions, seeking clarity. In that model, error is not something to eradicate but to interrogate.

Etienne Wenger (1998) described learning as participation in communities of practice, where ambiguity is negotiated collectively. Lev Vygotsky’s (1978) zone of proximal development is precisely the place where uncertainty lives: tasks the learner cannot yet do alone. In such spaces, error and approximation are not incidental but essential. Neil Postman and Charles Weingartner (1969) urged educators to cultivate “crap detectors”: habits of inquiry that thrive on questioning dubious claims. Hallucination could be the twenty-first-century “crap detector’s gym,” where students practise evaluating plausibility. Robert Ennis (1987) framed critical thinking as the ability to assess reasons and evidence; hallucinated outputs supply a steady stream of imperfect cases for such practice.

Contemporary metacognition research reinforces this. John Flavell (1979) described metacognition as knowing about one’s own knowing. John Dunlosky and Janet Metcalfe (2009) showed that when learners reflect on uncertainty, their learning accelerates. Anastasia Efklides (2006) argued that affective responses to uncertainty, frustration, curiosity, are part of metacognitive regulation. A classroom that openly engages AI hallucinations can cultivate precisely these skills: awareness of limits, strategies for verification, resilience in the face of ambiguity. This also aligns with Guy Claxton’s (2008) “learning power” framework, which positions uncertainty as a core dimension of lifelong learning. To navigate an unpredictable world, Claxton argued, students need to practise living with not-knowing. AI hallucinations, reframed pedagogically, become opportunities to build this epistemic muscle.

OpenAI named hallucination as structural error. But human sciences suggest it is more like structural possibility. It echoes how memory confabulates, how science advances through anomalies, how art invents, how learning thrives on desirable difficulties. Hallucination, then, is not simply a pathology of machines. It is an invitation to rethink pedagogy itself. Rather than demanding impossible certainty from AI, we might treat its hallucinations as provocations: imperfect stories through which students learn to discern, to critique, to imagine. To suppress

hallucination is to cling to an outdated vision of education as error-free transmission. To embrace it, carefully and critically, is to align classrooms with what we know of human flourishing: that uncertainty, if engaged wisely, is the ground on which new knowledge grows.

Pedagogical Implications – Uncertainty in the Classroom

Schools, as institutions, have long operated with a grammar of certainty. Larry Cuban (1993) described the “grammar of schooling” as the deeply embedded patterns, the timetable, the age-graded class, the focus on right answers, that persist despite reform. This grammar treats error as deficiency. In an exam, an error is a red mark. In a recitation, a wrong answer is an embarrassment. The system rewards closure and penalises ambiguity.

Yet history tells another story. John Dewey (1938) argued that genuine learning arises from “indeterminate situations,” when students encounter a problem that unsettles their assumptions. Seymour Papert (1980) insisted that the most powerful learning happens through debugging: confronting the uncertainty of why something doesn’t work and iterating toward understanding. Maria Montessori (1912) designed environments where trial and error was not punished but scaffolded. This tension defines modern classrooms: the institutional demand for certainty versus the psychological reality that uncertainty is the engine of growth. Into this contradiction now steps generative AI. Its hallucinations expose the fragility of the certainty grammar. A student can type a prompt and receive something fluent but flawed. How schools respond, whether with suppression or cultivation, will shape whether AI deepens the old grammar or disrupts it. The instinctive response to hallucination in schools has been policing. Universities warn students against fabricated citations (Flanagan & Shoichet, 2023). Examination boards fret about plagiarism detection (Selwyn, 2023). Ministries debate bans on generative AI in assessments. This stance mirrors what Michel Foucault (1975) described in *Discipline and Punish*: institutions exercise control by surveilling, policing, and punishing deviation.

But there is another path. Carol Dweck’s (2006) work on growth mindset, Angela Duckworth’s (2016) on grit, and Martin Seligman’s (2011) on resilience all converge on a key point: thriving learners are not those shielded from error but those trained to persist through it. James Paul Gee (2003) observed that video games, unlike schools, embrace failure as central to design: players fail repeatedly, but in low-stakes environments that encourage perseverance. If hallucinations are framed as failures to be policed, classrooms risk doubling down on the old grammar of schooling. But if hallucinations are reframed as practice grounds for resilience, they could teach students what Biesta (2010) calls the “beautiful risk of education”: the willingness to engage what cannot be fully controlled or predicted. Students would learn that credibility is earned not by avoiding error but by navigating it critically.

Designing Pedagogies of Uncertainty

So what would it mean to design classrooms where hallucination is not suppressed but harnessed? Several traditions in pedagogy offer clues. Jerome Bruner (1960) described discovery learning as a process of grappling with ambiguity. Lev Vygotsky (1978) placed learning in the zone of proximal development — the uncertain space just beyond current competence. Paulo Freire (1970) framed education as dialogical, rooted in questioning rather than depositing answers. bell hooks (1994) called for a pedagogy of freedom, in which uncertainty is not avoided but embraced as the condition of genuine encounter.

Building on these legacies, teachers might structure Hallucination Labs: assignments where the task is not to accept AI’s answer but to interrogate it. A history class could begin with a fabricated account of an event and challenge students to fact-check, cross-reference, and reconstruct. A science class could present a model’s speculative explanation and ask students to test it against real data. A literature class could treat an AI-generated misreading of a poem as the spark for discussion. Such practices align with Robert Ennis’s (1987) definition of critical thinking as “reasonable, reflective thinking focused on deciding what to believe or do.” They resonate with Neil Postman’s (1969) vision of inquiry education, where the role of the teacher is not to provide

the answer but to model the art of questioning. They echo Donald Schön's (1983) "reflection-in-action," where professionals learn by probing errors as occasions for deeper understanding.

Pedagogically, the shift is simple but radical: from correcting the machine to cultivating the student. The AI's hallucination is not a mistake to erase but a provocation to think. This reframing requires more than techniques. It demands cultivating what Lorraine Code (1991) called epistemic responsibility: the capacity to navigate knowledge claims ethically and critically. José Medina (2013) expanded this into the notion of epistemic virtues, humility, curiosity, courage, which enable agents to live with uncertainty without collapsing into cynicism or credulity.

In practice, this could mean teaching students to:

- Identify uncertainty: spotting when a claim may be speculative or unsupported.
- Interrogate sources: asking who benefits from a claim, what evidence underpins it, what biases shape it.
- Engage productively: seeing uncertainty not as paralysis but as invitation to inquiry.

David Perkins (1992) described this as "teaching for understanding," moving beyond rote knowledge to habits of mind. Guy Claxton (2008) called it "building learning power," embedding resilience, reflection, and resourcefulness. In the age of AI, these virtues are not optional extras; they are survival skills. Here, hallucination becomes a teacher. It constantly throws students into ambiguity, demanding discernment. The classroom becomes less a site of correct answers and more a workshop in epistemic agency. The teacher is not the guardian of truth but the curator of uncertainty. If the OpenAI paper was right, hallucination will never go away. Teachers cannot retrain the models. But they can retrain themselves and their students to interpret hallucination differently. The choice is stark: either schools cling to the inherited grammar of certainty, policing hallucinations as failures, or they rewrite the contract of learning to embrace uncertainty as integral.

The first option will preserve schooling as discipline: safe, predictable, but increasingly irrelevant in a world where uncertainty is the norm. The second will risk discomfort but cultivate resilience, creativity, and criticality. As Biesta (2013) reminds us, education is not about reducing risk but exposing students to the conditions of freedom. And as Maxine Greene (1995) urged, the task is to awaken imagination, to invite learners into the uncertain.

Hallucinations, then, are not obstacles. They are reminders. They remind us that knowledge is provisional, that learning is dialogical, that resilience is born in ambiguity. They remind us that classrooms, if they are to prepare students for the world they inhabit, must move from policing error to cultivating uncertainty. In this shift lies the revolutionary promise of AI in education: not as answer machine but as uncertainty partner, training the next generation not to fear the unknown but to live generatively within it.

Toward a (R)Evolutionary Pedagogy of AI

The Threshold Moment

When OpenAI released *Why Language Models Hallucinate* it did more than tidy up a technical debate (Kalai et al., 2025). It changed the cultural weather. For years, industry messaging implied that hallucinations were an inconvenience on the road to perfection: give us more data, more parameters, better guardrails, and the problem will fade. The paper punctured that dream. Hallucinations are structural, not incidental. They are the "cost of doing business" when your business is next-token prediction, where the model's imperative is fluency rather than truth. The system is built to sound right, not to be right.

The timing amplified the impact. In the same season, MIT researchers showed that habitual reliance on a large language model for intellectual work incurs what they called cognitive debt: thinner memory traces, less originality, lower persistence, what I earlier called "disappearing

thoughts” (Kos’myna et al., 2025; Hopkin, 2025). Meanwhile, public discussion swelled around mental-health harms and privacy fragility in mainstream chatbots; even their stewards acknowledged the need for new forms of oversight and care (Psychiatric Times, 2025; OpenAI, 2025). Three confessions in as many months: the human mind is tempted to offload; the machine, by design, will sometimes fabricate; the institutions behind them are themselves learning in public. It feels less like a “feature update” and more like a hinge of history.

Sociology gives us language for this hinge. Ulrich Beck called modernity a risk society, one in which the technologies that create prosperity also manufacture novel uncertainties we must learn to live with (Beck, 1992). Karl Weick’s work on sensemaking teaches that when environments become equivocal, organisations survive by organising ambiguity into action (Weick, 1995). Education has long pretended it is immune to such dynamics, insulating itself with syllabi and standardised tests. The OpenAI paper tears that illusion. If students offload their thinking to a system that cannot promise truth, then the naïve “transmission model” of schooling collapses: the teacher tells, the textbook confirms, the exam checks — and the machine serenely makes things up.

Yet the human sciences have also prepared us for this moment. John Dewey argued that inquiry begins in “indeterminate situations,” when experience fails to fit, and we are forced to investigate (Dewey, 1938). Jean Piaget made disequilibrium the engine of development (Piaget, 1977). Donald Schön showed that professional learning accelerates in the friction of error (Schön, 1983). Neuroscience goes further: dopamine systems prioritise prediction error, the gap between expectation and outcome, because that is where the organism learns the most (Schultz, Dayan & Montague, 1997). The paradox of our age is that we built machines whose errors we cannot abide, at the precise moment we most need a culture capable of learning from error.

So this chapter assumes the confession and asks the only interesting question left: If hallucinations are inevitable, what would it mean to design education that treats them not as pathology to police, but as pedagogy to harness? The answer, I will argue, requires a re-authoring of teacher identity, a redesign of classroom tasks and assessment, a reorientation of inspection and system accountability, attention to equity and workload, and a new literacy fit for a risk society. We will move from authority to dialogue; from accuracy-only to accuracy-and-discernment; from certainty culture to uncertainty competence (Freire, 1970; hooks, 1994; Wenger, 1998; Biesta, 2013).

From Authority to Dialogue (and the Teacher We Need Now)

The grammar of schooling, the repeated, taken-for-granted structures that persist despite reform, privileges certainty (Cuban, 1993). The teacher is the knower; correctness is the currency. Hallucinations, in that frame, are contaminants: they erode authority. But Freire called this a “banking” conception of education, a one-way deposit of knowledge, and proposed instead a dialogical practice: teachers and students co-investigate the world (Freire, 1970). bell hooks translated that into the ethics of the classroom: freedom as shared inquiry, not performance of deference (hooks, 1994). Lave and Wenger gave us the architecture: learning as participation in communities of practice, where ambiguity is negotiated in action (Lave & Wenger, 1991; Wenger, 1998).

Reframed through that lineage, hallucination ceases to be a personal affront to teacher credibility and becomes a generative interruption. Authority does not vanish; it evolves. The teacher’s authority is no longer founded on omniscience but on epistemic leadership: modelling scepticism without cynicism; scaffolding inquiry without abdication; converting plausible nonsense into a curriculum of discernment. In practice:

- English / Literature. The model misattributes a line to the wrong poet. A transmission teacher corrects it; a dialogical teacher turns it into a seminar: what internal evidence (lexicon, metre, intertext) points to the real author? What external evidence (editions, archives) counts? Students learn hermeneutics and habits of verification (Ricoeur, 1984; Fish, 1980; Wineburg, 2001).

- **History.** The model invents a treaty date. Instead of simply grading “wrong,” the teacher stages a micro-historiography: who says what, why, with what sources, for what purposes? Students practice sourcing, contextualisation and corroboration, the triad of disciplinary historical thinking (Wineburg, 2001).
- **Science.** The model offers an elegant but erroneous mechanism. Rather than correcting from the front, the teacher facilitates a quick-and-dirty experimental design to test claims. Students rehearse falsification as the engine of science (Popper, 1959; Osborne & Dillon, 2008).
- **Mathematics.** The model proposes a plausible-looking proof with a hidden gap. Students do error analysis, locating the break and rebuilding the argument. They learn that proof is a community practice of checking, not just a static product (Polya, 1945; Schoenfeld, 1985; Boaler, 2016).

This teacher stance requires specific forms of professional knowledge. Shulman called it pedagogical content knowledge: an understanding of how ideas are learned, including common misconceptions (Shulman, 1986). In an AI era, we need an allied construct: epistemic orchestration knowledge, the craft of using machine error to stage human learning. It blends metacognition (Flavell, 1979; Zimmerman, 2002), formative assessment (Black & Wiliam, 1998; Wiliam, 2011), and classroom discourse moves that surface uncertainty productively (Hattie, 2009; Mercer & Littleton, 2007). It also asks something moral: epistemic virtues, humility, curiosity, courage, the willingness to say “I don’t know,” to try, to test, to change one’s mind (Code, 1991; Medina, 2013).

How do we develop this teacher? Not by one-day workshops on “AI safety,” but through job-embedded professional learning communities where teachers design and trial Hallucination Labs, co-plan verification prompts, and study student work (Cochran-Smith & Lytle, 1999; Timperley et al., 2007). Leadership matters: schools with strong professional cultures buffer risk and sustain inquiry (Hargreaves & Fullan, 2012; Spillane, 2006). Systems matter too: inspection regimes that punish error will drive performative certainty; regimes that reward disciplined inquiry will grow epistemic courage (Ofsted, 2019; KHDA, 2024; OECD, 2024).

The pivot is subtle but profound: from the teacher as guarantor of correctness to the teacher as curator of uncertainty. In a world of inevitable hallucinations, this is not a downgrade; it’s a promotion.

Designing Classrooms for Uncertainty (tasks, assessment, inspection, equity, and the workload reality)

The rhetoric only matters if it lands in the timetable. What, concretely, does a pedagogy of uncertainty look like on Monday morning? Five design moves anchor it.

a) Design Move 1: Hallucination Labs

A Hallucination Lab is a deliberately engineered moment where an AI output becomes the object of study. The teacher curates a short, plausible, wrong answer, preferably one that reveals a typical pattern (fabricated citations, confident causal leaps, tidy but false mechanisms). Students then:

1. Diagnose the claim (what is being asserted?)
2. Triangulate sources (what evidence exists, and of what quality?)
3. Test or reconstruct (what experiment, derivation, close reading, or archival check would resolve it?)
4. Reflect (what heuristic misled us? what indicator will we notice next time?)

This workflow turns hallucination into a desirable difficulty (Bjork, 1994). It strengthens retrieval, deepens schema, and builds transfer. Crucially, it normalises productive discomfort: students learn that “I’m not sure yet” is a rational position, not a personal failure (Dweck, 2006; Dunlosky & Metcalfe, 2009).

b) Design Move 2: Verification as a Graded Outcome

If we only grade answers, we will incentivise blind reliance on fluent machines. Grade verification. In practice:

- Require students to append an Audit Trail to AI-assisted work: prompts used; drafts generated; sources checked; changes made; reasons why.
- Use triangulation rubrics that credit quality of sources, strength of warrants, and clarity about uncertainty bands (Ennis, 1987; Shanahan & Shanahan, 2008).
- Build micro-vivas: 3-minute oral defences where students explain how they vetted an AI claim.

This aligns assessment with the outcomes we say we value: discernment, not just correctness (Black & Wiliam, 1998; Wiliam, 2011).

c) Design Move 3: Collaborative Sense-Making

Uncertainty is easier to carry collectively (Weick, 1995). Structure group roles:

- Skeptic (probes assumptions),
- Verifier (checks sources / methods),
- Synthesiser (integrates perspectives),
- Communicator (explains the team’s judgment and its limits).

Rotate roles so dispositions distribute across the class (Vygotsky, 1978; Johnson & Johnson, 1999). Use protocols (think-pair-share; structured academic controversy) to ensure balanced airtime. The goal is not consensus, but warranted assent, “this is what we judge to be true, for now, on the weight of the evidence.”

d) Design Move 4: Subject-by-Subject Tooling

- History. Weekly “Source Trials” where an AI-generated paragraph is put on trial. Students prepare prosecution and defence using primary and secondary sources; the class judges credibility on agreed criteria (Wineburg, 2001).
- Science. “Mechanism Mondays”: the model proposes a mechanism; students design a low-cost test (Popper, 1959; NGSS Lead States, 2013).
- Mathematics. “Proof Clinics”: AI-written arguments with tiny gaps; students locate the breach, patch it, and write the metacommentary explaining why the gap is alluring (Polya, 1945; Schoenfeld, 1985; Boaler, 2016).
- Literature. “Attribution Games”: the model blends voices; students use stylistic fingerprints to separate them (Ricoeur, 1984; Fish, 1980).
- Civics / Media Studies. “Misinformation Drills”: students map claims through CRAAP/ HEART checks, reverse-image search, legislative databases, and local reporting (Kahne & Bowyer, 2017; Lewandowsky et al., 2012).

- Arts / Design. “Speculative Briefs”: students prompt an AI to over-generalise a style, then design work that “argues with the machine,” showing where genuine craft resists pattern (Csikszentmihalyi, 1996; Greene, 1995).

e) Design Move 5: Inspection, Policy, and the Workload Reality

A pedagogy that asks teachers to orchestrate uncertainty must not be smuggled in as extra. It needs system alignment.

- Inspection frameworks should reward epistemic outcomes: critical thinking, resilience, collaborative problem-solving (Ofsted, 2019; KHDA, 2024; OECD, 2024).
- Assessment policy should permit and expect AI use with verification, not pretend students won’t use it (UNESCO, 2023).
- Teacher workload must be protected: share task banks; co-design in PLCs; provide release time for planning; use lightweight tools for audit trails (Day & Gu, 2010; Hargreaves & Fullan, 2012).
- Equity must be explicit: scaffold verification protocols for novices; provide sentence frames for argumentation; design alternative representations for learners with SEND; ensure all students have access to devices and to quiet, supervised spaces for verification (Darling-Hammond, 2010; Dignath & Büttner, 2008).

A final practicality: transparency with parents. Families need to see that this is not laxity but rigour redefined: the move from “don’t use AI”, a performative impossibility, to “use AI, and show me your verification.” When parents see the audit trails and hear the micro-vivas, they recognise that their children are not marinating in confusion; they’re training discernment.

Global Perspectives, Critiques, and Counterpoints (with answers!)

a) Comparative Systems

Finland. With high teacher autonomy and phenomenon-based learning, Finland is well positioned to integrate Hallucination Labs as the front-end of inquiry cycles. The cultural tolerance for teacher judgment is an asset (Sahlberg, 2011).

Singapore. An exam-driven system might bristle at ambiguity, but the nation’s “Teach Less, Learn More” reforms and strong lesson-study culture could house verification-graded tasks in enrichment and capstones without destabilising high-stakes exams (Tan, 2013).

United Arab Emirates. With a policy appetite for innovation and bilingual classrooms, the UAE can frame uncertainty pedagogy as resilience, aligning with KHDA’s inspection emphases on innovation and wellbeing (KHDA, 2024).

United States / UK. Polarisation and misinformation make civic versions of this work urgent. Districts can embed misinformation drills into media literacy; Ofsted’s focus on a “rich, broad curriculum” already points toward disciplinary thinking rather than test prep (Kahne & Bowyer, 2017; Ofsted, 2019).

China / East Asia. Cultural attitudes toward face-saving may require careful staging of public error; small-group verification with teacher circulating privately can preserve dignity while growing skill (Hofstede, 2001).

Sub-Saharan Africa & South Asia. Infrastructure constraints suggest low-tech adaptations: printed AI outputs, community radio fact-checks, and offline verification kits can deliver the core pedagogy without heavy bandwidth (UNESCO, 2023). Equity first, always.

b) Critiques — and Credible Responses

“This invites relativism.”

Response: Only if we abandon standards. A pedagogy of uncertainty is standards-rich: explicit rubrics for warrant, source quality, and error-correction. It teaches Popper’s falsification, not “anything goes” (Popper, 1959; Ennis, 1987).

“Teachers are already overwhelmed.”

Response: True, so align the system. Provide shared task banks; timetable PLCs; give release time; assess verification instead of adding separate “critical thinking” projects (Day & Gu, 2010; Hargreaves & Fullan, 2012).

“Parents want certainty.”

Response: They want safety and competence. Show them audit trails and short orals; demonstrate that students can explain why something is true. This often increases parental confidence.

“We’ll widen gaps.”

Response: Only if we throw novices into ambiguity without scaffolds. Structured protocols, sentence frames, visual supports, and targeted feedback narrow gaps (Dignath & Büttner, 2008; Darling-Hammond, 2010). Remember Vygotsky: pitch in the Zone of Proximal Development (Vygotsky, 1978).

“Won’t students just use the AI to fake the verification?”

Response: That is why we pair written audit trails with live micro-vivas and in-class verification. The performance is the proof.

“Isn’t the real solution to fix the models?”

Response: Yes and no. We should demand safer, more transparent systems (Bender & Koller, 2020; Bommasani et al., 2021; Ji et al., 2023). But the history of technology says residual error is permanent. Education must cultivate human capacities for living intelligently with residual error.

A New Literacy and the Long Horizon (2030–2040): curriculum, teacher education, system design, and civic life

Claude Shannon defined information as the resolution of uncertainty (Shannon, 1948). Our mistake was to think we could eliminate uncertainty by outsourcing cognition to machines. The OpenAI confession requires a different bet: that the human project is to become good at uncertainty. Call it uncertainty literacy, a composite of knowledge, skills, and virtues that makes people competent in a risk society (Beck, 1992; OECD, 2024; UNESCO, 2023).

a) Competency Model

- Epistemic Humility: comfort with “I don’t know yet,” calibrated confidence (Code, 1991; Medina, 2013).
- Inquiry Craft: search, source evaluation, triangulation, replication; disciplinary ways of knowing (Wineburg, 2001; Shanahan & Shanahan, 2008).
- Metacognition: monitoring understanding; planning verification; regulating emotion under ambiguity (Flavell, 1979; Zimmerman, 2002; Dweck, 2006).
- Argumentation: constructing and critiquing claims; quantifying uncertainty bands (Osborne & Dillon, 2008; Ennis, 1987).
- Civic Discernment: misinformation resistance; deliberative dialogue; digital hygiene (Kahne & Bowyer, 2017; Lewandowsky et al., 2012).
- Creative Speculation: using imaginative conjecture responsibly — the arts’ contribution to knowledge (Csikszentmihalyi, 1996; Greene, 1995).

b) Curriculum Architecture (10–16)

- Weekly Hallucination Labs in core subjects (see Section 3).
- Verification Portfolio across the year: three artefacts per subject, each with audit trail and micro-viva.
- Civic Uncertainty Dialogues: termly cross-curricular roundtables on contested public claims (health, climate, economics).
- Capstone “Open Problem” (Y11/Y13): students select a real problem, use AI for ideation, and deliver a verified product with uncertainty analysis.

c) Assessment & Credentials

- Rubrics scoring warrant quality, handling of counter-evidence, clarity about uncertainty bands.
- On-Demand Orals (5–7 minutes) replacing slices of written exams in humanities and sciences (William, 2011).
- Credential Tags for “Uncertainty Literacy – Proficient/Advanced,” co-badged by universities/employers.

d) Teacher Education (Initial + Ongoing)

- Residencies focused on discourse moves that surface thinking and manage uncertainty (Shulman, 1986; Mercer & Littleton, 2007).
- Design Studios where teachers build/share Hallucination Labs; peer-video for feedback (Cochran-Smith & Lytle, 1999).
- Ethics & Wellbeing: protocols for student mental health, privacy, and bias in AI use (UNESCO, 2023).
- Leadership Pathways for “Epistemic Coaches” who support departments with verification assessment and parent engagement.

e) System Alignment (Policy / Inspection)

- Inspection indicators that look for verification evidence, not prohibition rhetoric (Ofsted, 2019; KHDA, 2024; OECD, 2024).
- Assessment regulators piloting orals and portfolio moderation at scale.
- Data policy mandating transparency over AI use in coursework; standardised audit-trail templates.
- Equity funding to guarantee access to devices, offline alternatives, and specialist support.

f) Scenarios 2030–2040

- 2030—Baseline Adoption: Hallucination Labs normalised, verification portfolios routine, short orals in national systems; teacher workload stable due to task banks; parental anxiety eased through transparency.
- 2035—Civic Impact: Students trained in uncertainty literacy measurably less susceptible to viral misinformation; community forums adopt school discussion protocols; local journalism partners with schools on verification projects (Kahne & Bowyer, 2017).
- 2040—Second-Order Effects: A culture that tolerates ambiguity without surrendering to relativism; innovation cycles faster because error is treated as data; teacher identity more intellectually attractive, since epistemic leadership is a high-status craft (Biesta, 2013; Greene, 1995).

The horizon here is not utopian. It is the plausible outcome of aligning what we assess with what we profess. If we say epistemic virtues matter but test only recall, we'll get recall. If we assess verification, argument, and calibrated confidence, we'll grow citizens capable of living well with uncertainty.

The Courage to Embrace

In medicine, a false negative can kill. In civic life, a fluent fabrication can poison a polity. In classrooms, a culture of certainty can make minds brittle. The OpenAI paper is not only an AI story; it is an education story. It dares us to stop outsourcing our courage, to accept that uncertainty is the native habitat of learning and to teach accordingly.

This is not the end of rigour. It is rigour, remembered. Popper's falsification. Dewey's inquiry. Piaget's disequilibrium. Schön's reflection-in-action. Wineburg's sourcing. Bjork's desirable difficulties. hooks' freedom. Freire's dialogue. Shannon's uncertainty. Biesta's beautiful risk (Popper, 1959; Dewey, 1938; Piaget, 1977; Schön, 1983; Wineburg, 2001; Bjork, 1994; hooks, 1994; Freire, 1970; Shannon, 1948; Biesta, 2013).

The machine will sometimes make things up. The mind will sometimes want to disappear. Our task, the teacher's task, the system's task, is to make those twin facts into a school that tells the truth about thinking: that knowledge is provisional, that confidence is earned, that inquiry is collective, that error is data, and that freedom is the courage to act without guarantees. If we can build such a school, hallucination becomes less a threat than a tutor: the friction that keeps the spark.

About the Author



Dr. Neil Hopkin is a globally recognised thought leader in international K-12 education, and serves as the Director of Education at Fortes Education.

His extensive academic background includes advising UK government bodies and spearheading significant educational initiatives, particularly with the EdTech, Early Years, Higher Education and Teacher Professional Development fields, equipping him with invaluable insights and expertise. As the head of Fortes' Academic Leadership Team, Dr. Hopkin is responsible for overseeing academic performance, operational efficiency, curriculum development, and staff professional development across Fortes Education institutions.

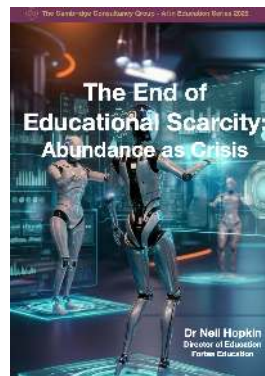
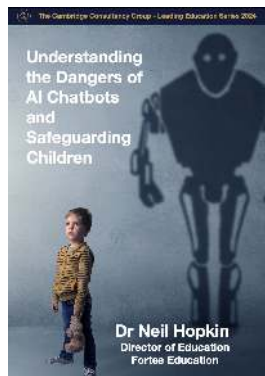
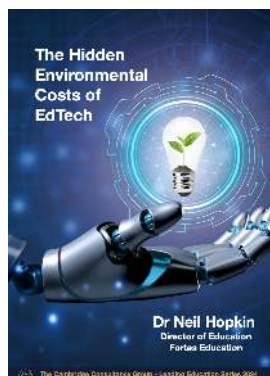
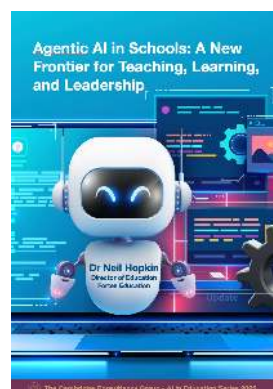
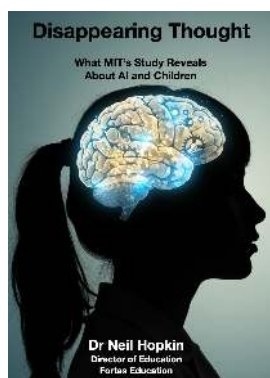


For more information contact Dr Neil Hopkin at:

www.sunmarke.com

www.risdubai.com

For other articles written by Dr Neil Hopkin, please visit his LinkedIn page at: www.linkedin.com/in/neilhoppin



Bibliography

- Amabile, T. (1996). *Creativity in Context*. Boulder, CO: Westview Press.
- Arendt, H. (1968). *Between Past and Future: Eight Exercises in Political Thought*. New York: Viking Press.
- Barron, F. (1963). *Creativity and Psychological Health*. Princeton, NJ: Van Nostrand.
- Beck, U. (1992). *Risk Society: Towards a New Modernity*. London: Sage.
- Bender, E. & Koller, A. (2020). "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5185–5198.
- Biesta, G. (2010). *Good Education in an Age of Measurement*. Boulder, CO: Paradigm.
- Biesta, G. (2013). *The Beautiful Risk of Education*. Boulder, CO: Paradigm.
- Bjork, R. (1994). "Memory and Metamemory Considerations in the Training of Human Beings." In Metcalfe, J. & Shimamura, A. (eds.), *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press, pp. 185–205.
- Bjork, R. & Bjork, E. (2011). "Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning." *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, pp. 56–64.
- Black, P. & Wiliam, D. (1998). "Assessment and Classroom Learning." *Assessment in Education*, 5(1), pp. 7–74.
- Boaler, J. (2016). *Mathematical Mindsets: Unleashing Students' Potential through Creative Math, Inspiring Messages and Innovative Teaching*. San Francisco, CA: Jossey-Bass.
- Bommasani, R. et al. (2021). *On the Opportunities and Risks of Foundation Models*. Stanford University: Center for Research on Foundation Models.
- Brashers, D. (2001). "Communication and Uncertainty Management." *Journal of Communication*, 51(3), pp. 477–497.
- Bruner, J. (1960). *The Process of Education*. Cambridge, MA: Harvard University Press.
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- Claxton, G. (2008). *What's the Point of School? Rediscovering the Heart of Education*. Oxford: Oneworld.
- Code, L. (1991). *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca, NY: Cornell University Press.
- Cochran-Smith, M. & Lytle, S. (1999). *Inquiry as Stance: Practitioner Research for the Next Generation*. New York: Teachers College Press.
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention*. New York: Harper Collins.
- Cuban, L. (1993). *How Teachers Taught: Constancy and Change in American Classrooms, 1890–1990*. New York: Teachers College Press.

Darling-Hammond, L. (2010). *The Flat World and Education: How America's Commitment to Equity Will Determine Our Future*. New York: Teachers College Press.

Day, C. & Gu, Q. (2010). *The New Lives of Teachers*. Abingdon: Routledge.

Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Holt.

Dignath, C. & Büttner, G. (2008). "Components of Fostering Self-Regulated Learning among Students: A Meta-Analysis." *Metacognition and Learning*, 3(3), pp. 231–264.

Duckworth, A. (2016). *Grit: The Power of Passion and Perseverance*. New York: Scribner.

Dunlosky, J. & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.

Ennis, R. (1987). "A Taxonomy of Critical Thinking Dispositions and Abilities." In Baron, J. & Sternberg, R. (eds.), *Teaching Thinking Skills: Theory and Practice*. New York: Freeman, pp. 9–26.

Fish, S. (1980). *Is There a Text in This Class? The Authority of Interpretive Communities*. Cambridge, MA: Harvard University Press.

Flanagan, A. & Shoichet, C. (2023). "Universities Warn Students over AI Fabrications." *The Chronicle of Higher Education*.

Flavell, J. (1979). "Metacognition and Cognitive Monitoring." *American Psychologist*, 34(10), pp. 906–911.

Foucault, M. (1975). *Discipline and Punish: The Birth of the Prison*. New York: Vintage.

Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Continuum.

Friston, K. (2010). "The Free-Energy Principle: A Unified Brain Theory?" *Nature Reviews Neuroscience*, 11, pp. 127–138.

Gazzaniga, M. (2011). *Who's in Charge? Free Will and the Science of the Brain*. New York: HarperCollins.

Gee, J. (2003). *What Video Games Have to Teach Us about Learning and Literacy*. New York: Palgrave Macmillan.

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of the Unconscious*. London: Penguin.

Greene, M. (1995). *Releasing the Imagination: Essays on Education, the Arts, and Social Change*. San Francisco, CA: Jossey-Bass.

Hargreaves, A. & Fullan, M. (2012). *Professional Capital: Transforming Teaching in Every School*. New York: Teachers College Press.

Hattie, J. (2009). *Visible Learning*. London: Routledge.

hooks, b. (1994). *Teaching to Transgress: Education as the Practice of Freedom*. New York: Routledge.

Hopkin, N. (2025). "Disappearing Thoughts." LinkedIn article, www.linkedin.com/in/neilhopkin

Ji, Z. et al. (2023). "Survey of Hallucination in Natural Language Generation." *ACM Computing Surveys*, 55(12).

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

- Johnson, D. & Johnson, R. (1999). *Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning*. Boston: Allyn & Bacon.
- Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), pp. 263–291.
- Kahne, J. & Bowyer, B. (2017). “Educating for Democracy in a Partisan Age.” *American Educational Research Journal*, 54(1), pp. 3–34.
- Kalai, A. et al. (2025). *Why Language Models Hallucinate*. OpenAI Research Paper.
- KHDA (2024). *Dubai Schools Inspection Framework*. Dubai: Knowledge and Human Development Authority.
- Kos'myna, N. et al. (2025). “Cognitive Offloading and Generative AI.” MIT Media Lab Working Paper.
- Kruglanski, A. & Webster, D. (1996). “Motivated Closing of the Mind: ‘Seizing’ and ‘Freezing.’” *Psychological Review*, 103(2), pp. 263–283.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lakatos, I. (1970). “Falsification and the Methodology of Scientific Research Programmes.” In Lakatos, I. & Musgrave, A. (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.
- Laidlaw, C. & Russell, S. (2021). “Value of Uncertainty in Preference Elicitation.” *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lewandowsky, S. et al. (2012). “Misinformation and Its Correction.” *Psychological Science in the Public Interest*, 13(3), pp. 106–131.
- Loftus, E. (2005). “Planting Misinformation in the Human Mind.” *Learning & Memory*, 12(4), pp. 361–366.
- Marcus, G. & Davis, E. (2022). *Rebooting AI*. New York: Vintage.
- Maynez, J. et al. (2020). “On Faithfulness and Factuality in Abstractive Summarization.” *Proceedings of ACL 2020*, pp. 1906–1919.
- Medina, J. (2013). *The Epistemology of Resistance*. Oxford: Oxford University Press.
- Mercer, N. & Littleton, K. (2007). *Dialogue and the Development of Children’s Thinking*. Abingdon: Routledge.
- Monosov, I. (2020). “Neural Circuits for Tracking and Resolving Uncertainty.” *Nature Reviews Neuroscience*, 21, pp. 579–593.
- Montessori, M. (1912). *The Montessori Method*. New York: Frederick Stokes.
- Nussbaum, M. (1997). *Cultivating Humanity: A Classical Defense of Reform in Liberal Education*. Cambridge, MA: Harvard University Press.
- OECD (2024). *Future of Education and Skills: OECD Learning Compass 2030*. Paris: OECD.
- Ofsted (2019). *Education Inspection Framework*. London: Ofsted.

- OpenAI (2025). Transparency Report. San Francisco: OpenAI.
- Osborne, J. & Dillon, J. (2008). *Science Education in Europe: Critical Reflections*. London: Nuffield Foundation.
- Palmer, P. (1998). *The Courage to Teach*. San Francisco, CA: Jossey-Bass.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books.
- Piaget, J. (1977). *The Development of Thought*. Oxford: Blackwell.
- Polya, G. (1945). *How to Solve It*. Princeton, NJ: Princeton University Press.
- Popper, K. (1959). *The Logic of Scientific Discovery*. London: Routledge.
- Postman, N. & Weingartner, C. (1969). *Teaching as a Subversive Activity*. New York: Delacorte.
- Psychiatric Times (2025). "ChatGPT and Mental Health." *Psychiatric Times*.
- Reason, J. (1990). *Human Error*. Cambridge: Cambridge University Press.
- Ricoeur, P. (1984). *Time and Narrative*, Vol. 1. Chicago: University of Chicago Press.
- Robinson, K. (2001). *Out of Our Minds: Learning to Be Creative*. Oxford: Capstone.
- Sahlberg, P. (2011). *Finnish Lessons*. New York: Teachers College Press.
- Schacter, D. (1999). *The Seven Sins of Memory*. Boston: Houghton Mifflin.
- Schoenfeld, A. (1985). *Mathematical Problem Solving*. Orlando, FL: Academic Press.
- Schön, D. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Schultz, W., Dayan, P. & Montague, P. (1997). "A Neural Substrate of Prediction and Reward." *Science*, 275(5306), pp. 1593–1599.
- Selwyn, N. (2023). *Education and Technology: Key Issues and Debates*. London: Bloomsbury.
- Seligman, M. (2011). *Flourish: A Visionary New Understanding of Happiness and Well-Being*. New York: Simon & Schuster.
- Shanahan, T. & Shanahan, C. (2008). "Teaching Disciplinary Literacy to Adolescents." *Harvard Educational Review*, 78(1), pp. 40–59.
- Shannon, C. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal*, 27, pp. 379–423.
- Shulman, L. (1986). "Those Who Understand: Knowledge Growth in Teaching." *Educational Researcher*, 15(2), pp. 4–14.
- Simonton, D. (2011). *Great Flicks: Scientific Studies of Cinematic Creativity and Aesthetics*. Oxford: Oxford University Press.
- Spillane, J. (2006). *Distributed Leadership*. San Francisco, CA: Jossey-Bass.
- Tan, C. (2013). *Learning from Singapore: The Power of Paradoxes*. Abingdon: Routledge.
- Timperley, H. et al. (2007). *Teacher Professional Learning and Development*. Brussels: International Academy of Education.

- Topel, S. (2023). "Uncertainty Strengthens Learning." *Nature Reviews Psychology*, 2, pp. 211–223.
- UNESCO (2023). *AI and Education: Guidance for Policy Makers*. Paris: UNESCO.
- Vygotsky, L. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Weick, K. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage.
- Wenger, E. (1998). *Communities of Practice*. Cambridge: Cambridge University Press.
- Wineburg, S. (2001). *Historical Thinking and Other Unnatural Acts*. Philadelphia: Temple University Press.
- William, D. (2011). *Embedded Formative Assessment*. Bloomington, IN: Solution Tree.
- Zhao, Y. (2012). *World Class Learners: Educating Creative and Entrepreneurial Students*. Thousand Oaks, CA: Corwin.
- Zimmerman, B. (2002). "Becoming a Self-Regulated Learner." *Theory Into Practice*, 41(2), pp. 64–70.