NEVER SURRENDER THE BATON

Why AI Makes Human Discernment More Important, Not Less

By Arthur Roark Leggett

September 21, 2026

Artificial intelligence is becoming the most remarkable orchestra humanity has ever assembled.

It can research, write, calculate, translate, synthesize, simulate, compare, challenge and search for patterns across vast volumes of information no individual human could reasonably absorb.

And it is not merely becoming more capable.

It is becoming ubiquitous.

Increasingly, AI will shape decisions, recommendations, analyses and the information people encounter, whether or not we consciously choose to engage it ourselves. Its output will flow through businesses, institutions, software and ordinary life.

Throughout this essay, I use AI broadly when discussing automated systems and institutional deployment. Examples involving language, hallucinations, reasoning and conversational collaboration refer primarily to generative AI, especially large language models.

That combination - greater capability and greater ubiquity - changes the human side of the equation.

As AI becomes more capable, human discernment becomes more valuable. As AI becomes more ubiquitous, human judgment becomes more consequential.

Those are not the same thing.

Discernment is the ability to distinguish.

Fact from inference. Signal from noise. Knowledge from conjecture. Genuine insight from polished mimicry. A hallucination from a claim supported by evidence. A conflation - two real things incorrectly fused together - from a genuine connection between them.

Judgment comes next.

Given what we have discerned, what should we believe? What should we reject? What deserves another look? When should we change direction? What should we ultimately do?

Discernment is the conductor's ear. Judgment is the conductor's baton.

Much of the conversation about artificial intelligence understandably concentrates on the orchestra: what increasingly capable machines can do and which human abilities they may eventually equal or surpass.

A harder question is what increasingly capable machines will require humans to become.

Beneath some of the ebullience surrounding AI lies a tempting inference: as artificial intelligence becomes more capable, human intelligence must become correspondingly less important.

I suspect something closer to the opposite may happen.

As cognitive production becomes cheaper, discernment becomes more valuable.

When polished arguments become abundant, recognizing a bad argument matters more.

When answers become abundant, determining which answers deserve belief matters more.

When hundreds of possibilities can be generated in mere seconds, deciding which possibility deserves pursuit matters more.

AI does not eliminate the conductor.

It places a vastly larger orchestra beneath the baton.

But even that metaphor contains dangerous assumptions: that the conductor is competent, the orchestra is cooperative and the score is trustworthy.

Real systems offer no such guarantees.

The conductor may mishear the music. A section may drift from the intended interpretation. Something unexpected may emerge. Occasionally that departure will be an error. Occasionally it may produce something better than anyone intended.

The metaphor has limits. Machines do not possess human agency in the same sense musicians do. But AI systems can still depart from instructions, constraints or expectations in ways their human operators did not anticipate.

That matters because the purpose of the baton is not blind obedience.

The baton coordinates agency; it does not abolish it.

A good conductor must do more than detect mistakes. The conductor must sometimes recognize that an unexpected note is not a failure at all, but the beginning of something worth hearing.

That is discernment.

And there is another complication.

Possessing the baton does not prove that someone deserves it.

Authority is not competence.

A conductor may have been properly selected, formally empowered and still fail to hear what is happening directly in front of them.

Which raises a question considerably more difficult than Who holds the baton?

Who gave it to them - and why?

Keeping a human "in the loop" does not, by itself, guarantee meaningful oversight. We still have to ask: What kind of human? Exercising what judgment? Operating under what incentives? With what authority? Accountable to whom?

Competence, however, does not solve another problem: the information itself can be wrong.

We tend to imagine AI failure as fabrication: the system invents a fact, a source, a quotation, a person or an event that does not exist. [1]

A hallucination.

That failure is serious, but conceptually simple. Something unsupported has been presented as though it were real.

A conflation is subtler.

The components may all be real. Two people exist. Two events happened. Two facts are correct. But the system joins them into a relationship that never existed.

Nothing has necessarily been invented.

Reality has been assembled incorrectly.

Because each component may survive individual verification, a conflation can escape superficial fact-checking.

There is another failure that may be harder still to recognize.

Sometimes a problem appears solved at the visible layer while the underlying error, cost, uncertainty or contradiction has merely moved outside the observer's frame. [2]

Call it displacement.

The answer may be coherent. The arithmetic may work. The citations may be legitimate. The facts may survive inspection.

And still something essential may be missing: an unexamined assumption, a transferred cost, an excluded variable or a downstream consequence never measured.

The system may have answered the question perfectly.

The question itself may have been wrong.

That is why verification cannot mean merely checking whether individual claims are true.

Sometimes every visible note is correct while the symphony remains wrong.

The harder question is:

What lies outside the frame?

Here the score analogy matters most. The score is not reality. It is a representation of reality.

Representations can be incomplete, outdated or distorted by poor measurement, selective evidence, inherited assumptions or the objectives of those who created them.

A system can execute that representation flawlessly and still produce the wrong result.

Discernment must therefore sometimes move one level higher.

Who wrote the score? What evidence shaped it? What was excluded? Which assumptions were treated as facts? What becomes visible when the frame is widened?

AI makes those questions more important because fluency can conceal the frame beneath the answer.

Weak reasoning once often revealed itself through weak execution.

AI can separate the two.

A questionable premise can produce excellent prose. A narrow frame can produce sophisticated analysis. A bad assumption can survive pages of internally consistent reasoning.

Better execution does not rescue a defective premise. It can conceal it.

Then comes volume.

AI can generate claims faster than humans can verify them. A polished analysis can appear in seconds while examining the assumptions beneath it may require hours.

The bottleneck shifts.

Generating information becomes cheap.

Verifying it remains expensive.

That imbalance creates verification overload.

The problem is not simply that humans may become lazy. The volume itself can make comprehensive verification impossible.

Eventually the temptation becomes obvious:

It looks right.

The sources are there.

The numbers seem reasonable.

Move on.

That is precisely where discernment is most needed - and where AI may make it hardest to practice.

Speed deepens the problem.

A system may take thousands of consequential actions before a human fully understands the first failure.

At some scale and speed, oversight performed only after the fact is not oversight.

It is archaeology.

Human judgment must therefore sometimes enter before execution: through limits, checkpoints, permissions, reversibility and the ability to stop the system when something departs from expectation.

That does not mean AI should be reduced to an obedient machine that never surprises us.

Quite the opposite.

Some of its greatest value may come from finding relationships, possibilities and alternatives a human never anticipated.

An unexpected result can be an error. A hallucination. A conflation. Drift.

Or discovery.

Discernment is knowing the difference.

Which brings us to the uncomfortable possibility that the human safeguard may itself be the problem.

What happens when the system exposes a flaw and the person with authority insists that the process continue?

That is the danger of the sanctioned incompetent.

The problem is not that someone seized authority.

The system gave it to them.

But incompetence is not the only danger. A person may be perfectly capable of seeing what is wrong and still allow it to continue because competence operates inside incentives.

An institution may reward speed over scrutiny, conformity over challenge, profit over caution, reputation over candor or completion over reconsideration.

That possibility is more troubling than simple incompetence.

An incompetent person may fail to recognize the problem.

A competent person operating under distorted incentives may recognize it perfectly - and choose not to stop.

Human judgment, then, cannot be separated from the system surrounding the person exercising it.

Who holds authority matters.

What rewards that person matters too.

Nor does useful collaboration require unquestioning machine obedience.

A sufficiently capable AI may identify evidence the human missed, uncover a contradiction the human introduced or challenge a premise the human has become invested in preserving.

Good.

That does not necessarily represent a failure of human control.

It may represent successful collaboration.

The human should not retain final authority because humans are necessarily smarter. That claim will become increasingly difficult to defend.

The stronger reason is responsibility.

Someone must remain answerable for what becomes action in the human world.

But responsibility becomes harder to locate as systems become more complex.

Real institutions rarely have a single conductor.

A corporation may distribute authority among executives, engineers, lawyers, boards, regulators, customers and automated systems. A hospital may divide consequential decisions among clinicians, administrators, insurers, protocols and software.

Each participant may exercise only part of the authority. Each may act rationally within a narrow role. Together they may produce an outcome none of them deliberately chose.

This is where complexity begins to obscure accountability.

Distributed authority can produce distributed responsibility. And distributed responsibility can quietly become no responsibility at all.

Everyone participated.

No one decided.

Everyone followed the process.

No one owns the outcome.

That should make us cautious about one of the most reassuring phrases in artificial intelligence: human oversight.

Oversight by whom?

With what information?

At what point?

With what authority?

And with what actual ability to intervene?

A person who can observe a system but cannot stop it is not necessarily exercising control. A committee in which responsibility is divided until no one feels accountable is not necessarily exercising judgment. And a human who merely accepts whatever the system recommends has not preserved meaningful human agency simply by remaining present.

The important question is not whether a human remains somewhere inside the machinery.

It is whether someone retains the capacity - and the willingness - to say:

No.
Stop.
Something is wrong.
Show me the evidence.
Widen the frame.
Reconsider the assumption.
Change direction.
Do not execute.

That is what the baton ultimately represents.

Not domination.

Not omniscience.

Not guaranteed competence.

Responsibility for judgment.

And even responsibility faces one final danger.

The longer we rely on AI, the more AI may begin to change the person relying on it.

Every tool changes its user. Calculators changed how we approach arithmetic. Search engines changed how we remember information. GPS changed how we navigate. AI may change how we think.

That is not necessarily bad. Good tools should reduce unnecessary labor.

But there is a difference between reducing labor and allowing a capacity to atrophy.

If we routinely delegate research, synthesis, writing, comparison, calculation and analysis, we may gradually practice those abilities less.

And that creates a paradox:

The technology that makes human discernment more valuable may simultaneously weaken the habits that develop it.

The danger is subtle because decline may initially look like increased productivity.

The work still gets done.

Perhaps faster.

Perhaps better.

The system compensates.

Eventually, however, a person may lose the ability to recognize when the output is wrong.

Then dependence has crossed into something more consequential.

We have not merely delegated production.

We have begun delegating the capacity required to judge the production.

That is the difference between delegation and abdication.

Delegation says: Help me think.

Abdication says: Think for me.

Delegation says: Show me possibilities.

Abdication says: Tell me what to believe.

Delegation preserves the baton.

Abdication hands it away.

There is another risk worth keeping in view.

As AI systems multiply, apparent diversity may conceal intellectual similarity. Different systems may produce different answers while drawing from overlapping information, assumptions, methods and sources.

Repetition can begin to resemble corroboration.

AI-generated output enters the world as an article, memorandum, report, webpage, database entry or social post. Later, another person - or another machine - encounters that output as information.

What began as output returns as input.

A hallucination can acquire descendants. A conflation can become familiar enough to acquire the appearance of truth.

That makes provenance - Where did this actually come from? - increasingly important.

So does intellectual independence.

The audience complicates matters too. People reward certainty over nuance, speed over reflection, reassurance over truth, outrage over proportion and elegance over accuracy. AI did not create those human tendencies. It can amplify them enormously.

Which returns us to a mistake this argument must avoid.

The answer is not to romanticize the human.

Humans hallucinate.

Humans conflate.

Humans follow bad incentives.

Humans surrender judgment to authority, popularity, ideology, institutions and habit.

Humans can be hubristic, careless, frightened, tribal and spectacularly wrong.

Putting a human in charge does not magically produce wisdom.

Nor should we pretend every human-AI system resembles a symphony orchestra. Some will have several conductors. Some may have no obvious conductor at all. Others may resemble jazz ensembles, markets, ecosystems or networks in which coordination emerges from many participants.

The metaphor should illuminate the problem.

It should not become another score we refuse to question.

But across those arrangements, one principle survives:

Capability does not eliminate responsibility.

That may be the deepest reason human judgment matters.

Not because humans will always know more.

Not because machines must always obey us.

Not because every unexpected machine action is a threat.

And certainly not because human judgment is infallible.

Human judgment matters because decisions eventually enter a human world of consequences.

Someone decides whether to trust the recommendation. Someone decides whether to deploy the system. Someone determines its permissions. Someone accepts or rejects its conclusions. Someone decides which risks are tolerable. Someone chooses when to stop.

Or at least someone should.

The challenge, then, is not simply keeping a human somewhere in the loop.

It is preserving the human capacity to discern, judge, question, challenge, redirect, stop - and accept responsibility - as AI becomes increasingly capable, ubiquitous and intertwined with our institutions.

That requires something from the machines: transparency where possible, boundaries, checkpoints, reversibility and the ability to expose uncertainty rather than conceal it.

But it requires something from us as well.

- Curiosity.
- Skepticism.
- Humility.
- Courage.

The discipline to verify.

The willingness to recalibrate.

And the ability to hear when something does not sound right even when every note appears to be in tune.

There is one final distinction.

Conviction is not obedience.

Conviction is recognizing what is true, necessary or right. Obedience is acting on what that recognition requires.

AI may sharpen the diagnosis, expose the contradiction, identify the risk or make the proper course of action painfully clear.

But clarity is not action.

We should not confuse seeing the problem with answering it.

Conviction without obedience is still abdication.

AI may become a remarkable collaborator.

It may challenge us.

Correct us.

Surprise us.

Occasionally it may go somewhere we did not direct it and discover something better.

We should want that.

The purpose of the conductor is not to suppress the orchestra. It is to help remarkable capability become meaningful performance.

The future of human-AI collaboration should therefore not be built around human supremacy or machine submission.

It should be built around responsible agency.

Let the orchestra become magnificent.

Let it play things we could never play alone.

Let it hear patterns beyond our hearing.

Let it challenge the score.

Let it challenge the conductor.

And when the unexpected note arrives, listen closely enough to determine whether it is a hallucination, a conflation, a warning, a mistake -

or the beginning of something beautiful.

Use the orchestra.
Trust it when trust has been earned.
Question it when the music demands questioning.
Stop it when stopping is necessary.

And never confuse extraordinary capability with the responsibility to decide what should be done with it.

Never surrender the baton.

REFERENCES

[1] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1), 2024. See also William H. Walters and Esther Isabelle Wilder, "Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT," Scientific Reports 13, 14045 (2023), https://doi.org/10.1038/s41598-023-41032-5.

[2] Dr. Samuel E. Gralla and Dr. Kunal Lobo, "Electromagnetic Scoot," arXiv:2112.01729 (2021). Its "scoot" concept inspired the displacement analogy; the AI application here is the author's.

ACKNOWLEDGMENT

An early conversation with Wes M. sharpened the questions of individual agency and who selects the conductor.