

WILEY

From hallucination to verification

Why trusted, responsible AI needs grounded knowledge

JOSH JARRETT • WILEY • AI FOR GOOD, GENEVA • JULY 2026



Answers you can trust?

The answers AI gives to questions about global health, climate science, and public policy are drawing 40% of the time, from Reddit

Where AI systems draw their citations

40.1%



#1 most cited source

26.3%



WIKIPEDIA

#2 most cited source

23.5%



#3 most cited source

Analysis of 150,000+ citations across ChatGPT, Perplexity, Gemini and Google AI Overviews (Semrush, June 2025)

The contamination is already here

Hallucinations in the scientific and policy record

South Africa AI Policy, April 2026

- Draft National AI Policy withdrawn 17 days after publication
- 6 of 67 cited sources were AI hallucinations: fake papers, fake authors, fabricated journals

The scientific record: Nature, 2025

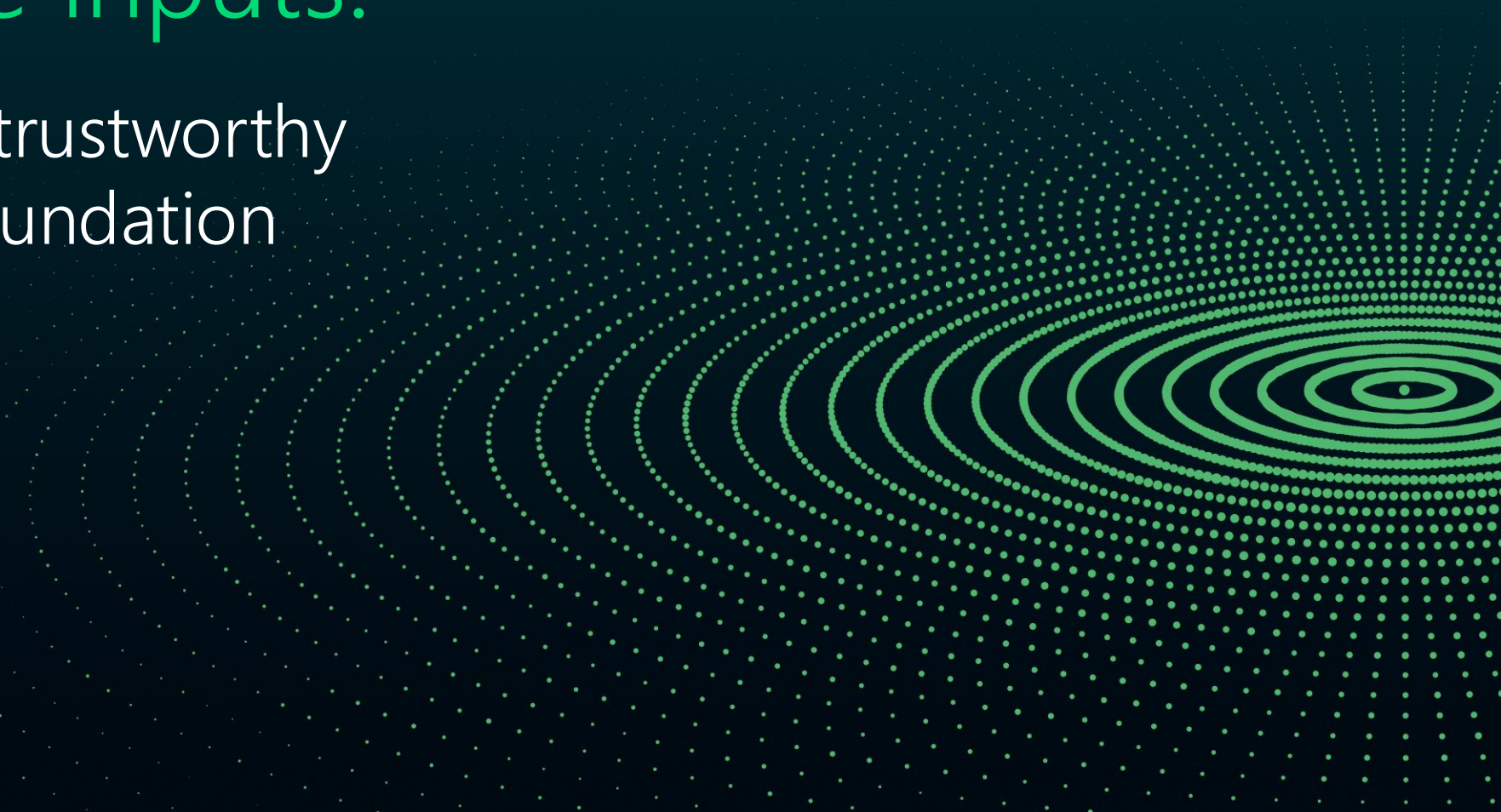
- 2.6% of all academic papers published in 2025 contain at least one hallucinated citation
- Up from 0.3% in 2024 — that is 110,000+ papers from 2025 with invalid references

We are not talking about chatbot errors.

We are talking about the contamination of the scientific record.

Responsible AI requires responsible inputs.

You cannot get trustworthy outputs if the foundation is unverified.



Five dimensions of responsible AI

TRUST

Confidence without verifiability makes AI dangerous in high-stakes contexts

General purpose chatbots are not regulated

FAIRNESS

Bias in, bias out

Peer-reviewed research from diverse global contexts is a fairness lever and an accuracy one

TRANSPARENCY

More than attaching a DOI

The difference between a citable finding and an assertion

HUMAN OVERSIGHT

Only works when there is something meaningful to review

Grounding AI in verified, attributable knowledge is vital

ENVIRONMENTAL SUSTAINABILITY

Every back-and-forth adds to an AI's carbon cost

RAG cuts hallucinations by over 40%, reducing correction cycles

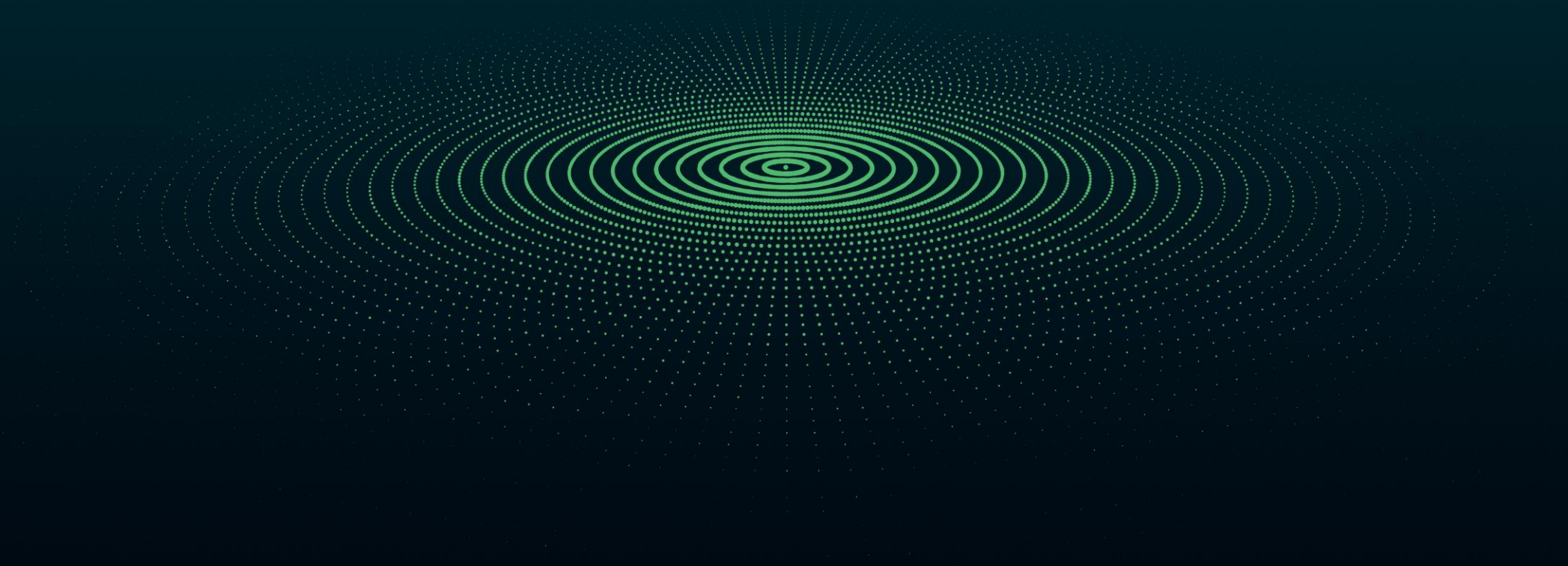
Trust: confidence is not the same as verifiability

- When outputs look credible, but sourcing is untraceable, the result is not trustworthy AI, it is **confident AI**
- Confidence without verifiability is precisely what makes AI dangerous in high-stakes contexts

ECRI 2026 Health Technology Hazards list

- Misuse of AI chatbots in healthcare ranked the **#1 health technology hazard**
- General-purpose chatbots are not regulated as medical devices and not validated for healthcare use

What the trust layer looks like in practice



The gap in current responsible AI frameworks

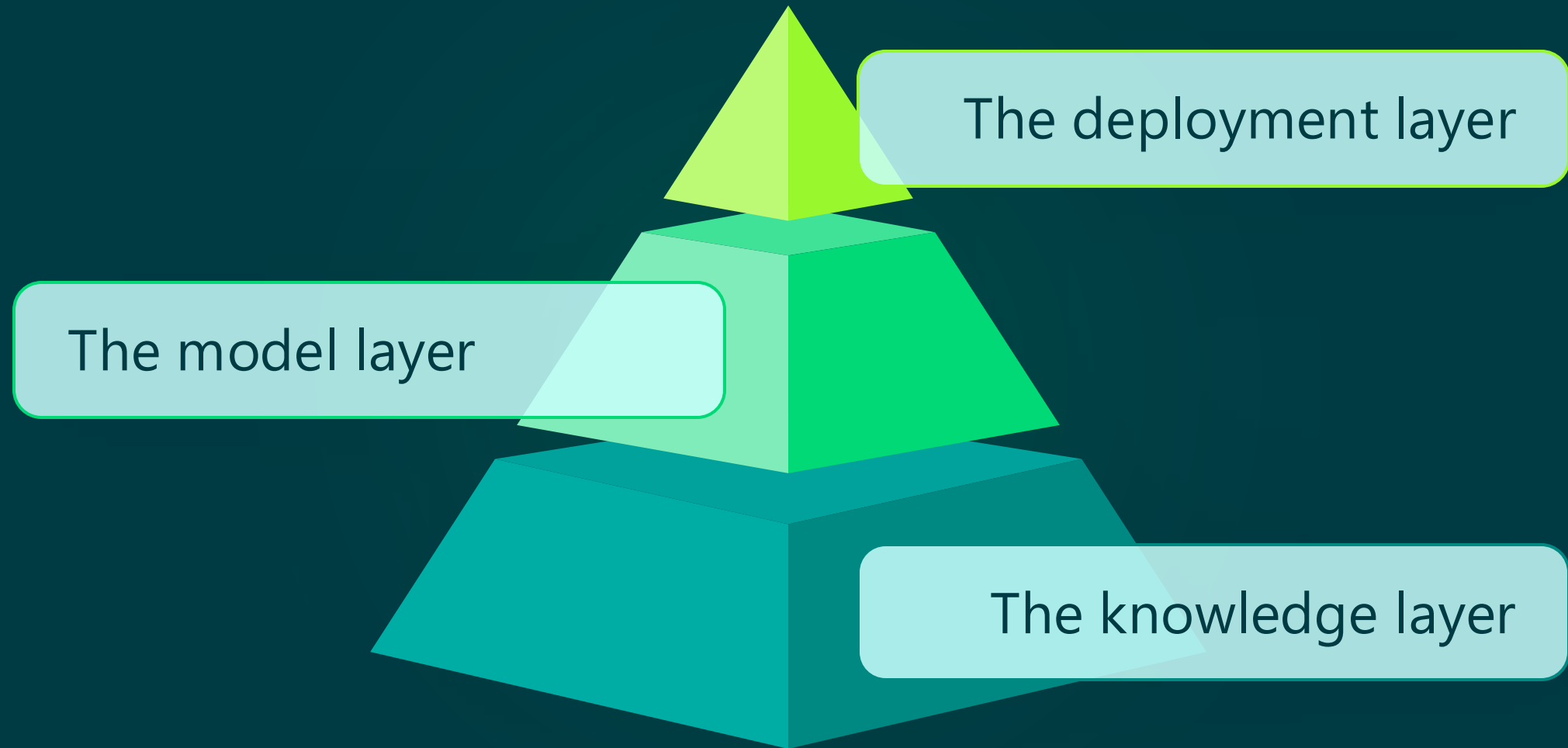
- The EU AI Act, NIST AI RMF, and UNESCO AI Recommendation govern what AI **does**, not what it **knows**
- None of these frameworks specify **source quality** as a requirement

That gap is
an opportunity.

We can ask the question now:

- What is this system built on?
- How was it sourced?
- Who is credited?

Three layers of trust



Two dimensions of knowledge quality

One infrastructure problem

Breadth

- Drawing from diverse, authoritative, current sources
- Not just what is most available or loudest online
- Represents researchers from all regions, languages, and disciplines

Depth

- What is inside the source, not just the abstract
- Citation tells you where something came from — full text determines what the system understood

Trust signals

Citations | Attribution | Retractions | Peer review status | Versioning | Corrections



Biorelate

Case study

What full-text access changes in AI for drug discovery vs. article abstract access only

63%

more novel biological
relationships per article
vs general open-access average

160K

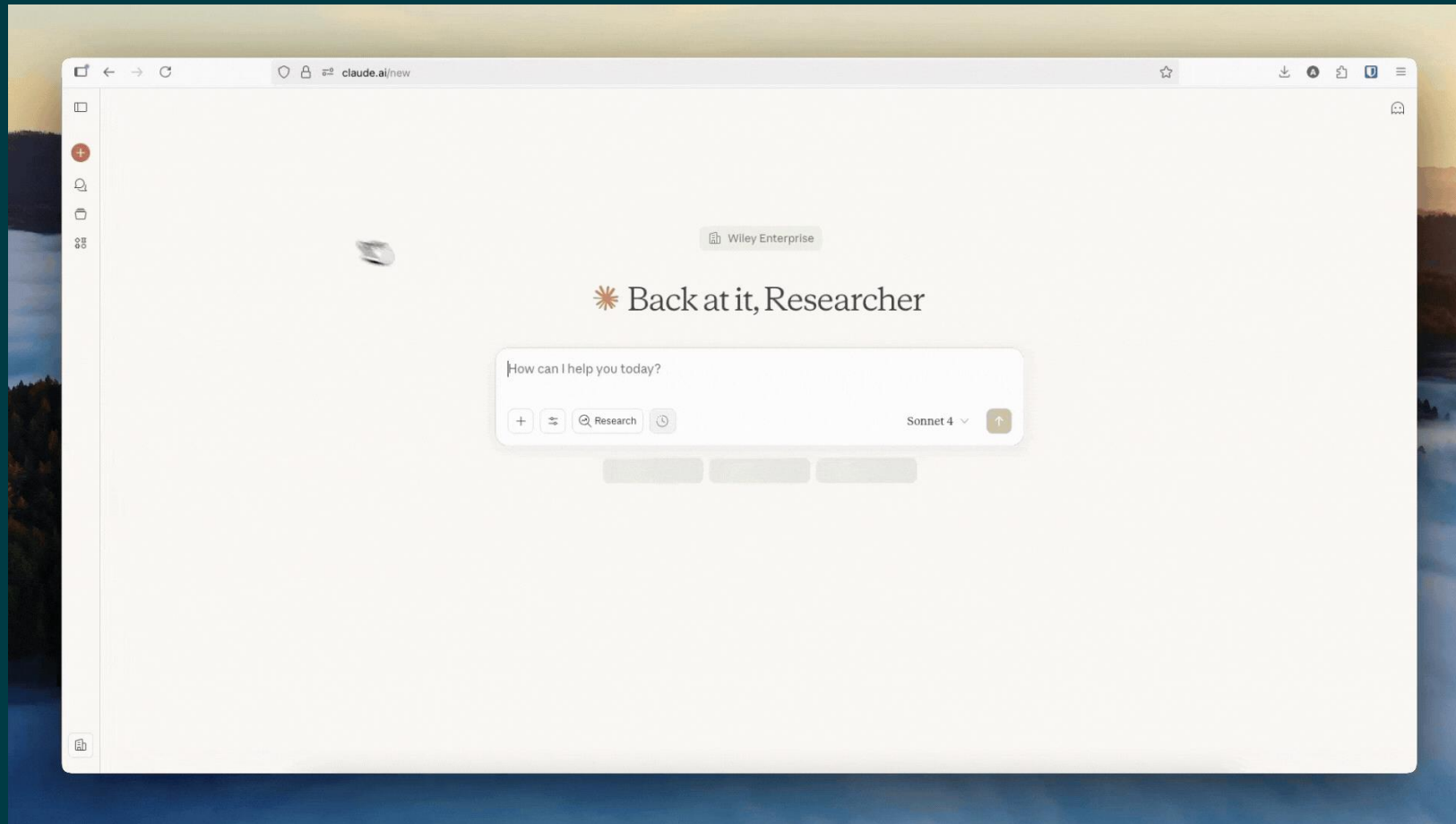
unique biological
relationships
surfaced nowhere else in PMC

8.4%

of Wiley's full-text
content in PMC
used in the proof of concept

Citations tell you where something comes from,
full-text content determines what the system understood.
Responsible AI needs both.

Demo: Scholar Gateway



The humanitarian stakes,
and what you can do



The trust crisis is already here

- We are entering a period where evidence is structurally harder to trust
- The volume is overwhelming. A new peer-reviewed paper is published roughly every ten seconds. A practicing specialist would need to read for 26 hours a day just to stay current in their own subspecialty.
- Synthetic content is cheap to generate and expensive to audit
- AI-generated text is flooding the research record, blurring the line between what was found and what was invented

"The progressive erosion of shared epistemic ground"

The slow collapse of the common evidence base that institutions, policy, and science depend on

Every AI system that cannot show its sources is not neutral.

It is part of the problem.

The trust stack

A framework for responsible AI in high-stakes contexts

Ground it

Ground any high-stakes AI application in authoritative, verified, full-text content.

Trace it

Preserve the scientific context, provenance, and chain of custody as content moves through AI systems.

Show it

Provide sources, citations, linkbacks, and reproducibility for every AI-assisted claim.

Evaluate it

Develop shared trust evaluation frameworks across the knowledge, model, and deployment layers.

What it looks like when we get it right



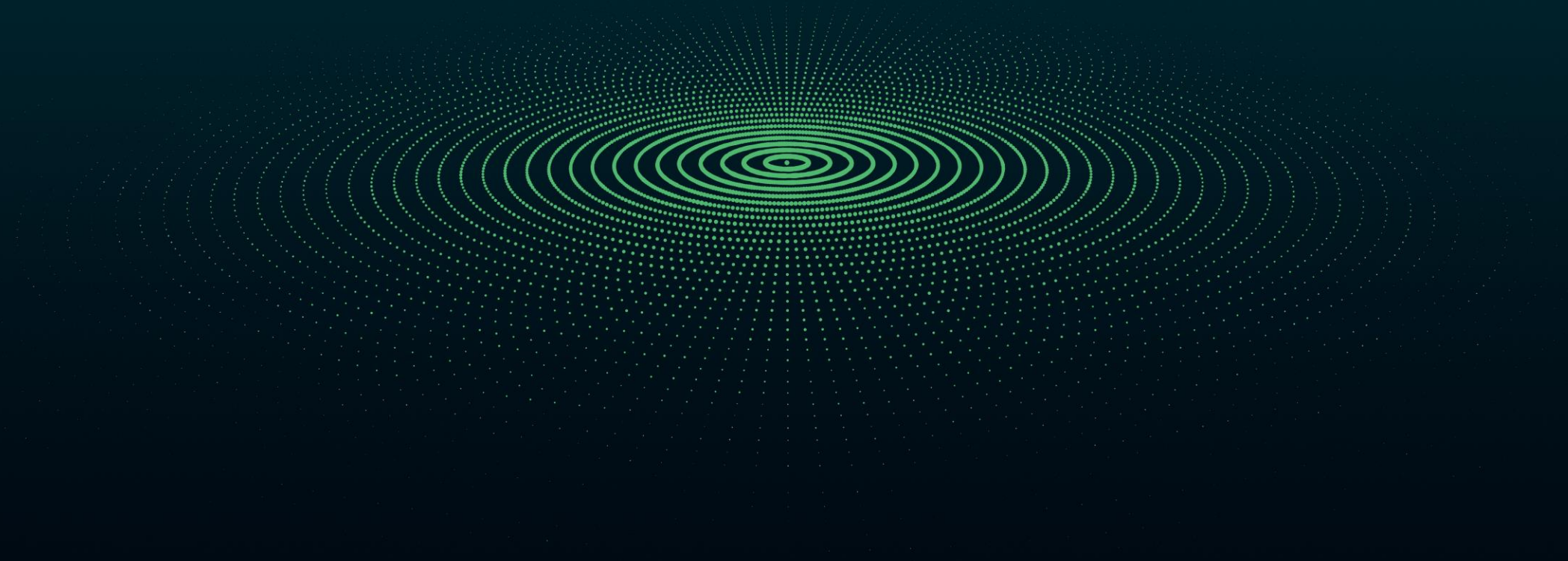
EARTH VIRTUAL EXPERT (EVE)

- Trained on Wiley's Earth science research, EVE makes satellite intelligence and Earth research accessible to policymakers
- Proper attribution on every output: Transparent about its sources, with an open-source roadmap
- AI serving global challenges, grounded in verified knowledge

WILEY + EUROPEAN SPACE

AGENCY + PI SCHOOL

Responsible AI starts with what the AI knows.



Visit us at booth 18

AI at Wiley



Find me on LinkedIn

