GW | ETHICS IN PUBLISHING 2026

Designing for What AI Can Know

Content Gaps, Metadata, and Integrity in Publishing Ecosystems



PRESENTER:
Stacy Masucci, Special Deputy, NYS DOS

AGENDA

Designing for What AI Can Know

01



Who Gets Published?

02



What Can AI Know?

03



How Do We Design for Integrity?

A REAL-WORLD EXAMPLE

The Research We Never See



South Africa

Helping researchers understand the publishing process.

- ✓ Proposals
- ✓ Submissions
- ✓ Publication

Understood the Process → Built Skills → Increased Representation → Monitored Geographic Trends



The real barrier wasn't the research.

It was knowing how to get through the door.



That experience taught me something I still believe today.

You can't address a gap you haven't learned to recognize.

Who Gets Through the Door?



The publishedrecord reflects both the **knowledge that exists** and the **opportunities to publish it.**

Both researchers have valuable knowledge.

Access and understanding of the publishing process influences opportunity.

The published record includes work that navigated the system.

Publishers can help create more equitable pathways.

Publishing equity is about opportunity, **not lower standards.**

A PUBLISHER'S PERSPECTIVE

Publishers have always shaped knowledge.



Every publishing decision helps shape the boundaries of the knowledge people can access.

THE BIGGER PICTURE

What Isn't There?



- Millions of books.
- Every subject imaginable.
- Every language.
- Every perspective.

...Or at least, that's what it appears to contain.

But how would you know what was missing?

Untranslated Works

An important work may never be translated.

Unsubmitted Research

A respected researcher may never submit their work.

Acquisition Decisions

A publisher may decide not to acquire a title due to a smaller market.

Licensing Limits

A book may exist but be unavailable on certain platforms.

Our publishing ecosystems are enormous, sophisticated, and increasingly accessible.

But they're not complete.

And the gaps aren't random—they're shaped by decisions made throughout the publishing process.

Where Content Gaps Come From:
The Knowledge Pipeline



Content gaps can emerge at many different points in the publishing process.

Our publishing ecosystems reflect not only what people know, but **what the publishing industry has been able and willing to make available.**

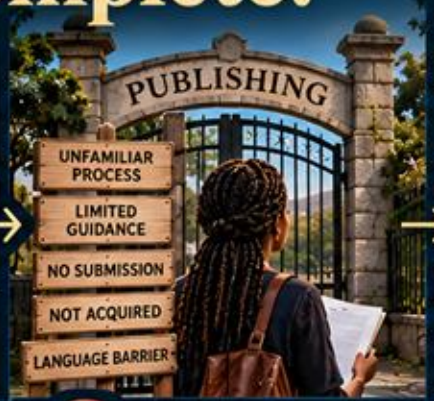


And increasingly, **artificial intelligence is drawing on those ecosystems.**

AI Can Make Incomplete Information Sound Complete:



Valuable research is conducted.



The research never reaches publication.



The published record is incomplete.



AI produces a polished, confident response based on the sources it can access.



The answer may be accurate based on the available sources. **But it can still be incomplete** when important research is missing.

INCOMPLETE ≠ WRONG

INCOMPLETE
AI can accurately summarize the evidence it has retrieved, even if that evidence doesn't represent the full range of relevant knowledge.

HALLUCINATION
AI generates information that isn't adequately supported by sources.

When relevant knowledge is absent from the sources an AI system can access—or difficult for that system to retrieve—the quality and completeness of its answers can suffer.

Absence Becomes Distortion:

When knowledge is missing, understanding can be incomplete.

Content Gap

Important research is never published, is not acquired, is not translated, or is difficult to access.

Limited Evidence

AI searches the available published record and finds only a partial view of the topic.

EUROPE
NORTH AMERICA
GLOBAL PERSPECTIVE
SELECT REGIONS
AFRICA
LATIN AMERICA
OTHER REGIONS

Incomplete Answer

AI produces a polished, confident response based on the sources it can access—but key perspectives may be missing.

AI Response
- Summarizes available studies
- References respected sources
- Presents a convincing answer

Publishers are not responsible for every limitation of AI.

But **we do have a role in shaping the knowledge** those systems may rely on.

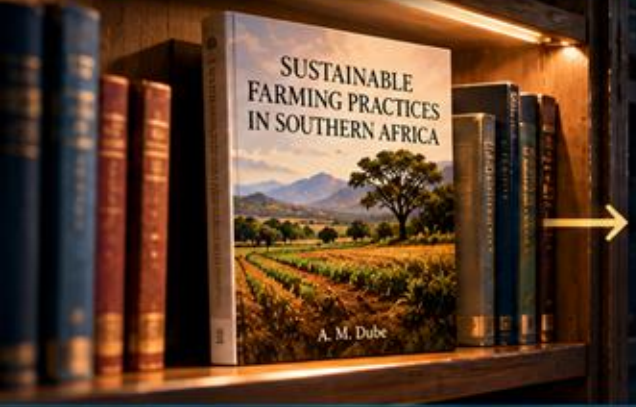
We may not build the algorithms, but we help determine which books and research become available. And that gives us **influence.**

With that influence comes **responsibility.**

We cannot assume that what has been published represents everything worth knowing.

A Book Can Exist and Still Be Invisible:

Even when a book is published, its **discoverability** depends on metadata.



METADATA

Title:	Sustainable Farming Practices in Southern Africa
Author:	A. M. Dube
Subject:	Agriculture – General
Description:	A study of farming methods.
Keywords:	farming, agriculture, crops
Category:	Agriculture

SEARCH RESULTS

climate resilient agriculture Africa

This book exists. But it doesn't appear in the results.

A valuable book is published.
The research is conducted, edited, and brought to market.

→

Metadata is incomplete or inaccurate.
The subject, description, keywords, or category don't fully reflect the content.

→

The book is hard to discover.
It may not appear in traditional searches or AI-powered systems.

→

The people who need it may never find it.
Important audiences can't access the knowledge.

→

A book can exist and still be invisible.
Discoverability is not just a marketing issue—it's an access issue.



Access affects **whose ideas enter the conversation.**



Integrity Doesn't Happen by Accident

Integrity by Design

From good intentions to **intentional systems.**

It's easy to identify problems in publishing. It's much harder to develop solutions that organizations can realistically implement.

A MORE EQUITABLE KNOWLEDGE FUTURE

DIVERSE VOICES
GLOBAL PERSPECTIVES
ACCESS
DISCOVERY
INCLUSIVE FUTURES

INTEGRITY BY DESIGN

PEOPLE
PROCESSES
POLICIES
PRACTICE
IMPACT

FOUR QUESTIONS FOR PUBLISHERS

1

WHO GETS A CHANCE?

How can we remove barriers and expand opportunities for diverse authors and researchers to participate?

2

WHAT DO WE ACQUIRE?

Are our acquisition decisions reflecting a broad range of perspectives, subjects, regions, and languages?

3

CAN IT BE FOUND?

Are we using high-quality, consistent metadata to ensure content is discoverable across today's and tomorrow's systems?

4

CAN IT BE ACCESSIBLE?

Are we considering licensing, formats, translations, and platform availability so content can be used widely?

Integrity isn't an afterthought.
It's **a deliberate part** of how we work.

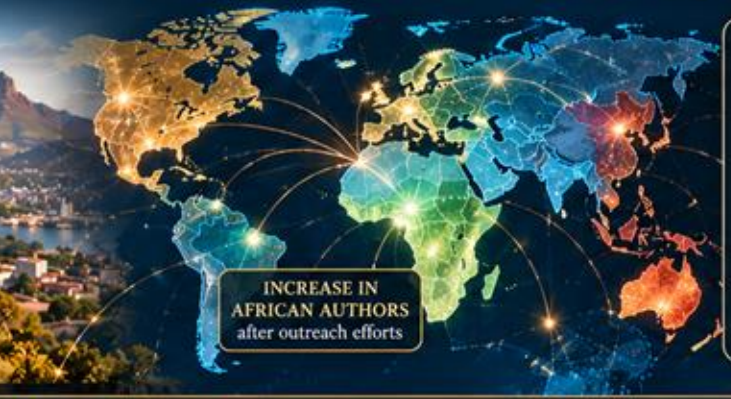
We may not be the people building the algorithms, but we help determine which books and research become available. And that gives us **influence—and responsibility.**

Who Is Missing?

And how would we know?

It's easy to **make assumptions** about how representative our publishing program is. But without data, we may not recognize who is consistently **underrepresented**.









Monitoring geographic representation doesn't tell us why a gap exists, but it gives us **a starting point** to move from assumptions to evidence—and from evidence to **more intentional decisions**.

As we see patterns, we can ask better questions:

Is the issue a lack of submissions? → Are potential authors unfamiliar with our publishing process? → Are we reaching the right research communities? → Could targeted education, outreach, or partnerships help?

From aspiration to accountability — **a more inclusive and representative publishing future.**

What Is Missing?

Looking beyond what performs to understand unmet demand and overlooked knowledge.

We analyze what sells. What readers purchase. What content performs well. But we should also be examining **unmet demand**—so our acquisition decisions are **more informed** and we understand the **consequences** of what we choose not to acquire.

KEY QUESTIONS TO ASK

- What are readers searching for that they **cannot find?**
- What subjects are consistently **underserved?**
- What works are unavailable because of **licensing limitations?**
- What perspectives are **missing** from our collections?



Commercial success matters.
But it's not the whole story.

Unmet demand reveals opportunities
to better serve readers and research communities.

More informed acquisitions
lead to stronger, more inclusive collections.

Understanding what we don't acquire helps us see the **real-world impact** of our choices.

What Is Invisible?

A published book can still be hard to find if it isn't accurately represented and organized.

We need metadata and classification systems that reflect how real people search, including different terminology used by different communities, so knowledge is **genuinely accessible** to everyone.



METADATA

- Subject classifications
- Descriptions
- Keywords
- Indexing practices

DIFFERENT TERMS, SAME IDEA

Water management	↔	Water resources
Water security	↔	Hydrological resilience
Rural water supply	↔	Smallholder irrigation
Community water systems	↔	Local water solutions
Indigenous water knowledge	↔	Traditional ecological knowledge

A book is published.
The research exists and is available.

→

Metadata is incomplete or inconsistent.
Subject terms, descriptions, or keywords don't fully reflect the content.

→

It becomes harder to find.
Search systems may not recognize alternative terminology or perspectives.

→

Important audiences may miss it.
The people who need the knowledge may never see it.

→

Knowledge remains less accessible.
Invisible in search results doesn't mean the book doesn't exist.

Good metadata is more than an operational requirement—it's part of making published knowledge **genuinely accessible.**

What Does AI Overlook?

AI can provide helpful and accurate answers — but its responses are only as complete as the content it can access.

As we adopt AI-powered search and recommendation tools, we need to evaluate them **intentionally** — testing what they retrieve, what they miss, and how **transparent** they are about their sources and **limitations.**



HOW CAN WE EVALUATE AI?

- Test retrieval across different languages, regions, and perspectives.
- Examine whether certain content is consistently overlooked.
- Make sources clear and accessible to users.
- Be transparent about limitations when working from a limited collection.

AI provides convincing answers.
It can summarize available research accurately.

→

It relies on the content it can access.
If relevant works are missing or hard to retrieve, they won't be included.

→

Important research may be overlooked.
Especially from underrepresented regions, languages, or perspectives.

→

Users may not recognize the gaps.
The answer can be accurate, but still incomplete.

→

Trustworthy AI acknowledges limits.
It helps people understand the evidence behind an answer—and the limits of that evidence.

Trustworthy technology isn't technology that always sounds certain.
It's technology that helps people understand the evidence — and the limits of that evidence.



From Outreach to Accountability

A real example from South Africa shows what's possible when we move from **good intentions** to **intentional systems.**

You can't address a gap you haven't learned to recognize. And recognizing a gap is only the beginning. The next step is **building a process** that allows you to **keep paying attention.**

1 Recognize the barrier

Researchers needed guidance on proposals and submissions.



2 Take action

Provide publishing education and outreach.



3 Observe the outcome

More African authors represented in publications.



4 Build accountability

Monitor publishing geography over time.



What started with **outreach** led to a **more inclusive scholarly record.**

That's what **integrity by design** looks like — not just addressing a gap, but building a process to keep recognizing and learning from it.

Who Isn't in the Room?

Whose research **aren't we seeing?**
Whose perspectives **aren't represented?**
Whose work exists but remains **difficult to discover?**
And what might **future generations misunderstand** because that knowledge never became part of the record?

We won't eliminate every barrier.
We won't publish every worthwhile book.
And we won't create perfectly complete knowledge systems.

But we can be **more intentional about the systems we build.**



Make publishing more accessible.

Identify gaps in our collections.

Monitor the geographic representation of our authors.

Improve how content is described and discovered.

Ensure AI-enabled systems are more **transparent** about the knowledge they rely on.

Integrity in publishing isn't only about protecting the quality of what we publish—
it's about taking responsibility for the knowledge we make possible to discover.