

Society of Indexers Conference 2025: 'The Heart and Mind of Indexing'

**Adapting with Agility: how indexing  
societies are responding to AI challenges**  
Society of Indexers update

Tanya Izzard  
20 September 2025

#SIConf25

# Overview

- Terminology
- AI and publishing update
- How can indexers and indexing societies respond?
  - research
  - policy statements

# Terminology

- LLMs/generative AI: AI tools like chatbots or based on chatbots that use predictive algorithms to respond to prompts, drawing on the data contained within a Large Language Model (LLM)
- Machine Learning (ML): the development and study of statistical algorithms that can learn from data and to perform tasks without explicit instructions, for example the analysis of medical data (generative AI tools also make use of ML)
- Artificial General Intelligence (AGI): AI that would match or surpass human capabilities across virtually all cognitive tasks. Not yet achieved, much discussed.

# AI in the media

- ‘jagged edge’ reporting: some AI tools perform well and improve productivity, others are less helpful, but we hear more about the first part
- vagueness about which type of AI is meant in reporting
- boosterism (“AI will solve all our problems!”), doomerism (“AI is the end of the world as we know it”), cakeism (“we can have all the benefits without any cost or risk”)

# AI use in publishing

## Benefits

- Content creation – research, drafting, analysis
- Editing – applying in-house standards, suggesting/making corrections
- Marketing – data analysis, generation of materials
- Metadata creation
- General administrative streamlining and cost reduction

## Risks

- Copyright and security of intellectual property
- Cognitive debt and deskilling
- Impact of AI on SEO and of AI shopping tools (Amazon's Rufus)
- Environmental and reputational
- Need for investment from staff and staff training
- Investment without productivity gains

# LLM/Generative AI tools and indexing

- LLM/gen AI tools
  - chatbots (ChatGPT, Claude, Gemini, ...)
  - dedicated indexing tools built on chatbots
    - Tandem Index
    - Index Studio
    - LeanPub Index AI
- How they perform
  - tend to under-index with significant omissions
  - introduce errors, inaccuracies and hallucinations
  - cannot pick up implicit topics
  - cannot create navigable index structure
  - cannot empathise or imagine the needs of the reader – a solely human skill

# Indexing societies' responses: research

## **ASI White Paper on LLM-generated Book Indexes abstract:**

“Book indexes allow readers to quickly access the information they seek. Necessary criteria for a book index include completeness (an index must guide the reader to all indexable information in the book), navigability (an index must guide the reader to subtopics and related topics), and accuracy (an index should not contain false or inaccurate information and should reflect the author’s terminology). We investigated whether artificial intelligence (AI) based on large language models (LLMs) is capable of providing indexes that meet these criteria and found those indexes fall far short. At this time, AI cannot replace professional book indexers and we are doubtful that it will be able to do so soon, or even at all.” (<https://digital-publications-indexing.org/index.php/statement-on-ai-in-indexing/>)

# Indexing societies' responses: research

Ongoing research:

- selection and human indexing of sample texts for experiments with AI tools
- systematic testing of new releases of chatbots
- systematic testing of dedicated indexing tools
- some costs involved in this – indexer time, cost of access to chatbots/tools
- publication of results – white papers, *The Indexer*, blog posts
- co-ordinated activities within and between indexing societies desirable to avoid duplication of effort and establish a consistent position

# Indexing societies' responses: statements

## DPI-SIG Statement on the Role of AI in Indexing (as of July 1, 2025)

“Our current assessment is that using large language model AIs (LLMs) such as ChatGPT to generate indexes does not produce results that come anywhere near meeting our standards for excellence in Indexing. We also advise that documents not be uploaded or shared with LLMs without explicit client permission.

**These LLMs fail at the indexer's primary task: to ensure readers find needed information.** Tests have shown that LLMs typically underindex book-length works, do not provide adequate structure or cross-references, and can insert false information (hallucinations) into the index.\* Each of these is problematic:

- **Severe under-indexing**, with as few as 20-40% of the access points of human-generated indexes, can prevent readers from locating desired information and mislead them into thinking that omitted information isn't in the book at all
- **Absent/near-absent index structure**, especially cross-references, prevents the reader from effectively navigating to subtopics and related topics while misrepresenting the focus of the book
- **Hallucinations** (including invented page references and even wholly nonexistent topics) waste the reader's time and break the trust between the reader and the book

Future developments in AI may bring improvements, but at present we conclude **the human brain of a professional indexer is still the best tool** for analyzing, writing, and editing an index in full awareness of context as per quality standards.” (<https://digital-publications-indexing.org/index.php/statement-on-ai-in-indexing/>)

# Indexing societies' responses: statements

- Should we develop a statement for the SI website?
  - emphasising value of human indexing
  - promoting quality indexes
  - developing indexing skills
- If so, what would it cover?
  - limitations of current AI tools
  - reinforcement of the value of human indexing
  - signposting to more information about indexing and indexers

# Individual indexers' responses

- Statements for websites/SI profiles:
  - commitment to not using AI tools for indexing and/or in your work as an indexer (if that's the case)
  - links to institutional statements, research
  - DPI SIG has a suggested template but you can make it your own
  - you may already be committed to this in some of your client contracts
- To bear in mind:
  - a lot of software and apps now come with AI tools built in (Adobe, Microsoft Copilot etc)
  - if you are committing to not using AI tools at all, you need to turn these off, as well as avoiding AI search
  - be intentional about your engagement with AI and clear about your usage

“I don't use AI tools or any form of automated indexing in my work. It's just my brain with your book, working out what to select and how to present it in the index.”

# Summary

- AI tools are increasingly prevalent, but accurate information about use cases and genuine productivity gains can be hard to identify
- Indexing societies can respond by creating research-based, authoritative information about AI and index and promoting this, as part of their mission to promote quality indexes
- This might include clear public statements on the limitations of AI for indexing, and this work needs to be coordinated across societies
- Individual indexers might choose to make their own statements about AI use