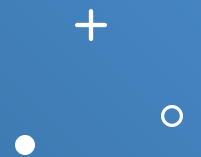




# How AI is being used in publishing

Presentation to Society of Indexers, 18 September 2024

George Walkley, Outside Context Ltd



# Introduction

- +
  - — Independent consultant, board director, doctoral researcher
  - 27+ years experience in the book business, across bookselling, publishing, consulting
  - Digital transformation experience at Hachette
  - Worked with IPG to develop first independently accredited, publishing specific training on Generative AI and publishing
  - 240+ publishers in 10+ countries: Big 5, independent trade publishers, university presses, specialists
  - Aggregate perspective on how AI is, and could be, used



# Publisher Overview

- + ●
  - — Almost every publisher experimenting with AI to some degree (many licensing)
  - Increasing integration with publishing application stack
  - Competitive advantage
  - Stakeholder concerns
  - Guardrails around tools, copyright content
  - Focus on process, marketing, ancillary content rather than core content (copyright questions)
  - Output quality
  - Immediate ROI for many publishers, especially smaller teams



# Publishing Use Cases

- +
  - - — Ideation/longlisting ideas
    - Summaries/many-to-few
    - Metadata creation
    - Content repurposing/reformatting
    - Content creation (ancillary materials)
    - Content translation (within limits)
    - Audio and video
    - Automation of tasks (VBA, Python, JS)
    - Stock imagery (Firefly)
    - Custom GPTs

In general, non-specialist tasks



# Pace of Change and Capability Overhang



- Fierce competition between AI companies
- Significant improvements in models with human-level performance across range of benchmarks (Stanford Institute for Human-Centered AI)
- ChatGPT o1-Preview: enhanced reasoning capabilities
- Absent any further development, 5-10 years to understand systems we have today (Ethan Mollick)
- Learning more from experimentation than documentation
- Balancing waiting for certainty versus near-term benefits



# Accuracy and Provenance

- +
- 
- 

- Probabilistic nature of models and non-exhaustive training data mean hallucination is inevitable, even with mitigations such as RAG
- AI does not have intrinsic knowledge, judgement, nuance or consistency
- Bias to high resource languages in training data sets
- Types of error: large and obvious versus more subtle (Ethan Mollick)
- Identifying AI-generated content easier in other media formats (audio, images, video) compared to text
- Standards for disclosure

- +
- 
-

# Copyright and Fair Use

- +
  - 
  -

- No doubt that copyrighted material has been used extensively in training models, without explicit authorization
- Fair Use/economic benefit argued by AI companies
- Litigation from authors and publishers
- Significant increase in licensing activity in 2024, but stakeholder concerns (Taylor & Francis, Wiley)
- Legal position is unlikely to be fully clear for some time, and there are short term commercial pressures to make the most of AI
- Fair Use finding is not impossible

- +
  - 
  -

# Resourcing and Tasks



- Impact of AI on economy: most people pessimistic about personal prospects (Pew Research Center)
- Particular concern from creative freelancers (editors, translators, illustrators)
- Large corporations also pessimistic about AI investment: “too much spend, too little benefit” (Goldman Sachs GMR)
- Thus far AI in publishing not delivering on wholesale disruption, but incremental gains in productivity
- Time to revisit the Aggregation of Margin Gains





# Conclusions



1. AI is hard to avoid and there is a measurable ROI from getting it right
2. AI capability will improve significantly over time, but increasing training costs favour general versus specific models
3. For most people, impact will be in terms of productivity and communications rather than specialist tasks
4. Human and AI, rather than human versus AI
5. Assume what has gone before has already been trained on
6. Clarity about how work will be used in future





# Thank you



hello@outsidecontext.co.uk



georgewlkly



/georgewalkley

