



# AI and Publishing: Towards a Balanced and Ethical Approach

Presentation to BookNet Canada, 10 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



# Overview

- + ●
- 

## 1. Key Issues:

- Pace of change and capability overhang
- Copyright and Fair Use
- Accuracy and provenance
- Environmental impact
- Application integration
- Resourcing and tasks

## 2. Example Publishing Use Cases

## 3. Principles for Balanced and Ethical AI Use

## 4. Questions

- + ●
-

# 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 Index)
- 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



# 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

# Accuracy and Provenance

- +
- 
- 

- Probabilistic nature of models and non-exhaustive training data mean hallucination is inevitable, but risks can be mitigated
- Issue of biases to high resource languages in training data sets
- Types of error: large and obvious versus more subtle (Ethan Mollick)
- Not all publishing use cases have same risk profile
- High stakes use cases: RAG
- Identifying AI-generated content easier in other media formats (audio, images, video) compared to text
- Standards for disclosure

- +
- 
-

# Application Integration

- Early LLMs were single purpose tools requiring specific intention
- Generative AI now built (or being built) into every aspect of the publishing software stack: Office, Workspace, Creative Suite, GitHub, productivity applications, bibliographic software, ERP...
- Not just application-level but operating system-level integration: Microsoft Copilot, Apple Intelligence on iOS
- AI is now pretty hard to avoid...

# Environmental Impact



- Significant environmental impacts of AI include increased electricity and water usage, comparable to national level (HBR)
- But AI also drives efficiencies and trade-offs elsewhere: e.g. 550% rise in computing workloads led to only 6% rise in data center resource consumption (Informa)
- Hard for publishers as end-users to accurately assess own impact
- But that's also true for most Scope 3 emissions: any economic activity requires judgement on impact and trade-offs





# 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 promise of disruption, but often delivering incremental gains in productivity, in particular around ideation and automation
- Time to revisit the Aggregation of Margin Gains



# Publishing Use Cases

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

Impact is immediate and measurable,  
*especially in smaller publishers*



# Principles for Balanced and Ethical AI Use



1. AI is hard to avoid and there is a considerable ROI from getting it right
2. Human-in-the-loop always
3. Map issues and impacts across entire value chain, especially around prepublication content
4. Identify higher risk use cases (e.g. content origination, asserting copyright, diverse perspectives)
5. Play to AI's strengths: ideation, drafting, repurposing, automation
6. Involve all stakeholder groups in formulating policy, and communicate clearly
7. Use AI resources mindfully
8. Start with smaller, measurable projects





Thank you  
Any questions?



hello@outsidecontext.co.uk



georgewlkly



/georgewalkley

