



# PEOPLE PLUS MACHINES

The role of Artificial Intelligence in Publishing

# Contents

Executive Summary	3
1 Introduction	11
1.1 How we define publishing for this study	12
1.2 Current evidence on the use of AI in the UK creative industries	13
1.3 Defining AI in the context of publishing	14
1.4 Our methodology and evidence base	15
1.5 Report structure	18
2 A taxonomy of AI in publishing	19
2.1 Taxonomy structure	20
2.2 The overall taxonomy	23
3 Evidence on the role of AI in publishing	31
3.1 Investment levels	32
3.2 What is the nature of AI investment?	33
4 The barriers to AI investment	44
4.1 Large AI-active publisher barriers	45
4.2 SME publisher AI barriers	51
4.3 Comparing AI investment barriers for large AI-active and SME publishers	52
5 The importance of AI for publishing in the future and recommendations	53
5.1 The future importance of AI for publishing	53
5.2 The potential for AI to transform publishers' organisations	54
5.3 The potential for AI to transform publishers' competitive landscape	56
6 Recommendations for supporting future AI investment by publishers	57
Annex A Stakeholder interview guide	60
Annex B The industry survey	62
Annex C Literature review bibliography and sources	64



# Executive summary

The UK government's Industrial Strategy places heavy emphasis and public investment on Artificial Intelligence (AI) as a driver of business innovation and future productivity growth. The Publishers Association asked Frontier Economics to provide an evidence-based assessment of the role of AI in the publishing sector.

To the best of our knowledge, this report is the first systematic analysis of AI in publishing for the UK. Based on sector interviews, case studies and an industry-wide survey, we develop a use case taxonomy of AI for publishing, and evidence on publishers' behaviours and attitudes around AI investment.

## KEY INSIGHTS AND POLICY RECOMMENDATIONS

- **AI is being applied throughout the value chain by some academic, education and consumer publishers to drive benefits for their organisation** (such as improved IP protection, content discoverability, market prediction and other strategic insights) and for their customers (by conducting routine search and summarisation tasks and generating new insights, AI is freeing up researchers, authors, teachers and consumers to focus on value-add or creative tasks).
- **Overall, the majority of publishers, irrespective of size and sector, consider that AI will be important over the next five years.** Of the publishers we surveyed that are already investing in AI, the majority have realised benefits, and all expected to do so within the next few years.
- **AI investment in the sector has just begun. Larger publishers are leading the drive.** Most investment began within the last three years. Publishers use small internal AI research teams, and collaborations with AI-tech start-ups and university researchers.
- **To increase publishers' AI investment levels and drive increased consumer benefits, the sector must overcome a number of investment barriers**, including a lack of technical AI skills and general awareness of the benefits of AI, difficulties implementing AI solutions with existing IT infrastructures and across siloed work streams, the importance of legal certainty regarding UK Intellectual Property (IP) Law, and for smaller publishers, the significant upfront investment costs associated with AI research and implementation.
- **We recommend that the industry and government in the UK work together to raise awareness of key AI investment issues**, promote engagement and identify policy and other solutions to address them. Key policy areas include: ensuring legal certainty regarding UK IP law; promoting collaboration between publishers, AI-focused SMEs and academia; and helping SME publishers access AI finance and skills.

The remainder of this summary pulls together key insights from extensive desk research and primary evidence gathered through conversations with publishers, an industry survey and case studies on:

- **Defining AI** (and a key aspect of AI, Machine Learning (ML)) in the context of the publishing sector;
- **Developing a ‘use case’ taxonomy** of AI in publishing linking applications of AI to user and wider benefits;
- **Identifying the barriers** to AI investment and adoption faced by publishers;
- **Publishing and AI in the future**, particularly around how the use of and benefits from AI in publishing are likely to evolve; and
- **Recommendations** for how policy can best support publishing in overcoming investment barriers and driving benefits.

## Defining publishing and AI

The scope of this report includes academic, education and consumer publishing. Our analysis does not cover newspaper, catalogue and magazine publishing, or software and computer games.

A key finding is that there is no single definition of AI: the term AI is used broadly to cover a range of applications. Our working definition for this study was developed in consultation with publishers, recognising that the scope of AI is constantly changing and can be interpreted differently depending on the specific application.

## DEFINING ARTIFICIAL INTELLIGENCE FOR THE PUBLISHING SECTOR

Artificial Intelligence is a broad term that covers multiple technologies which enable computers to sense, comprehend, act and learn. All current uses of AI in the publishing industry involve Machine or Deep Learning - either alone, or in combination with other technologies such as Natural Language Processing (NLP), Voice Recognition or Computer Vision.

## Our approach and evidence base

The evidence and analysis for this study was generated in four steps (see below) and relied on extensive primary and secondary evidence gathered over the period November 2019 to March 2020.



The primary evidence was gathered through:

- a series of seven **telephone interviews** with stakeholder publishers;
- an **online survey** sent to all 135 members of the Publishers Association which generated 31 responses, including around two-thirds of the Association's larger members; and
- three in-depth **case studies** constructed through telephone interviews and desk research.

The secondary evidence was compiled through an in-depth literature review covering AI use cases and the relevant regulatory issues concerning AI investment for the publishing industry.






# An AI taxonomy for publishing

Our taxonomy is structured around ‘value chain – technology – benefits’. This reflects how most publishers think about AI investment decisions in their organisation.



We learnt that most investment decisions are benefit driven – ‘Can we address this issue, be more effective with AI, deliver significant benefits to our customers?’. In some organisations, technology-fit is also important – ‘Can we apply AI to drive benefits within the current organisational structure / with these resources?’.

The role of AI in publishing can largely be represented by the summary taxonomy below. The main report provides more detail including nuances of the taxonomy for the specific academic, consumer and educational publishing sub-sectors.

Publishing value chain 	Technology 	Application (not yet deployed)/unique to strand	Internal benefits 	Wider benefits
Content acquisition by authors	NLP/TDM ML & DL	Content extraction and text summarisation Content creation	Automate repetitious tasks Automate knowledge work Deliver strategic insight Innovate faster	More efficient use of resources/time-saving
Content acquisition by publishers	NLP ML & DL	Identifying market trends	Deliver strategic insight Cost reduction	
Content and product development	NLP ML & DL	Language checks Plagiarism checks Copyright infringement	Automate repetitious tasks Automate knowledge work Better risk mitigation Cost reduction	More product and service innovation leading to higher quality and choice
Marketing, promotion, sales	NLP ML & DL	Recommendation engines Demand forecasting	Automate repetitious tasks Automate knowledge work Deliver strategic insight Innovate faster Change the way customers engage	Improved operational efficiency
Manufacturing, fulfilment, warehousing	Decision management systems, ML, AI optimised hardware/robotics	Stock management	Deliver strategic insight Change the way customers engage Cost reduction	An improved customer experience – product and service satisfaction
Finance*	NLP, ML	Automatic invoicing	Cost reduction	
Customer service*	NLP ML & DL Speech recognition	Chatbots	Change the way customers engage Cost reduction	

**Key:** NLP – Natural Language Processing TDM – Text and Data Mining ML – Machine Learning DL – Deep Learning

The taxonomy highlights that the AI technology focus for publishing is a combination of text and image recognition technologies with machine and deep learning. AI can be applied in a wide range of ways throughout the publishing supply chain to generate benefits both to publishers and consumers – for example, time-saving benefits, speeding up, improving the quality of and providing additional insights for innovation activities, improving operational efficiency and cost savings, and improving the customer experience.

These initial publishing benefits should generate subsequent wider benefits for the economy as a whole, including companies' improved ability to compete through innovation and efficiency gains, increased academic attainment levels, lower cost and faster medical break-throughs generating patient well-being benefits, and the well-being benefits associated with improved customer satisfaction.

## Evidence on AI investment activity

The online survey provides a current snapshot on the nature of AI investment by publishers. The responses we received from large publishers (employing at least 250 FTE staff worldwide) appear to be representative of larger UK publishers. We received only one response from a smaller publisher that was already investing in AI. Therefore our survey insights focus on larger, 'AI-active' publishers. More detailed research is needed for smaller publishers to fully understand their AI-related behaviours.

Our findings suggest that though a few publishers were investing in AI as far back as 2014, most of those currently investing first did so in 2017. Total investment in AI is increasing year-on-year. The majority of publishers currently investing in AI have small internal teams of around 1 to 5 staff focused on AI research, with around half of these based in the UK.

**Almost all of the large publishers who responded said they were currently using AI in their organisation, or were exploring how it could be used.** In terms of our taxonomy, the responses reveal that:

- **AI is being applied throughout the publishing value chain**, with perhaps less emphasis to date on customer service. Large publishers are most commonly using AI to acquire and develop new content (45% of large AI-active publishers). In the near future this may change to AI being most commonly applied to provide marketing and sales solutions: three-quarters of large publishers are already using or plan to introduce marketing and sales solutions within the next two years.
- **The most common application of AI in publishing at present is content classification** (for example, using meta data tagging to improve the discoverability of their content in conjunction with recommendation engines). Other common applications include using AI to identify market trends and support recommendation platforms.

- **Two-thirds of these large AI-active publishers have already realised benefits from their AI investments.** Within two years, all large AI-active publishers expected to have realised AI benefits. The most commonly cited realised benefits include improved IP protection and risk management (for example from an improved ability to detect plagiarism), increased competitive advantage and improved strategic insight.

The case studies provided further evidence of how AI can improve the discoverability of relevant academic content through metadata tagging and the development of research engines, helping support higher quality research outputs. We also show how AI can develop educational content to deliver improved insights to students and time saving benefits to teachers, enabling them to spend more time with their students.

One case study also illustrated how AI has been used to generate new publishing content, applying ML algorithms to the vast catalogue of scientific research. This application could unleash significant benefits in the field of academic research, helping rapidly select, categorise and summarise significant quantities of text data both more quickly and, potentially, more accurately than can be done by human researchers. We also found some examples of AI being used to generate consumer content. While stakeholders consider it unlikely that publishers will be marketing AI-generated fiction novels in the near future, there is considerable innovation activity in this area. More generally, these examples of AI generated content in publishing illustrate the potential for AI to increase writers' ability to create, through to freeing them up from routine research tasks and/or rapidly providing them with new scientific or creative insights.

## AI investment barriers

The survey evidence suggests that **a lack of AI-related skills** and **difficulties applying AI solutions with existing IT infrastructure** are the most common AI investment barriers faced by large AI-active publishers.

We found evidence of publishers overcoming skills and technology barriers through **collaboration with external research organisations**: half of the large publishers we surveyed on this issue consider themselves largely dependent on external expertise for acquiring AI skills and technology.

**Lack of awareness of the potential benefits of AI** also appears to be a significant barrier for large AI-active publishers. We understand this may be linked to a lack of scientific knowledge by investment decision-makers, and may also be attributable to the fast rate of technological change currently seen for AI. The evidence we gathered suggests that overcoming this awareness barrier could also unlock several other issues identified by some stakeholders, such as organisational barriers (e.g. willingness to invest in AI solutions that cut across organisational silos, requiring buy-in from multiple decision makers).

Though the direct evidence we have on barriers facing SME publishers is more limited, we did find that **large upfront costs associated with researching and implementing AI solutions may be prohibitive for smaller publishers**.



## Publishing and AI in the future

We asked publishers about how important they felt AI would be for the industry in the next five years. On a scale between 0 and 100, **publishers on average scored the future importance of AI for the sector at 69**. This suggests a high degree of future importance.

Our survey evidence suggests that AI has the potential to significantly transform publishers' organisations. **One in six publishers expect to experience significant transformation** and all large publishers expect AI to have at least a small impact. It may take some time for this transformation to begin. **Around half of large publishers expect transformation to begin within three years** and the remainder within ten years. **The scale and speed of transformation is expected to be slower by SME publishers**, however: one-third do not expect AI to transform their organisation, and the majority expect that any transformation will not occur within the next five years.

AI may also have a **significant impact on publishers' competitive environment**. Almost two-thirds of publishers (and 80% of large, AI-active publishers) expect to be competing with different types of organisations in the future as a result of AI. Our evidence identified AI-tech start-ups, and established tech companies turning to publishing-related AI applications, as potential future competitors.

## Recommendations

Investment in AI by publishers has the potential to generate significant benefits for the UK publishing industry and the wider economy, including productivity enhancement. AI technologies do not necessarily represent a threat to the current labour force in publishing – rather, they could be significant enablers of creativity, allowing authors, researchers and other creatives to invest less time in routine tasks and more time in generating new content. Many publishers believe these benefits are obtainable within the next two to three years, with the right levels of investment, though SMEs may be further behind.

Our research has identified three areas of concern regarding AI in publishing, which could represent key barriers. We recommend that the industry and government in the UK work together to raise awareness of the issues, promote engagement and identify policy and other solutions to address them.

### 1

**Ensuring legal certainty with regard to UK Intellectual Property law. In particular, our analysis identified the following specific concerns:**

- Rights issues associated with using text and data mining (TDM) to conduct research, given that publishers hold the rights to significant amounts of data that are required for TDM to be applied.

- Copyright protection and legal responsibility for AI-generated works, in particular around concerns that AI technologies that summarise content or make editorial decisions could introduce bias, misrepresent information or infringe copyright in underlying content.
- Patenting of AI-created content, as patents are typically awarded to incentivise human creativity yet there is no clear guidance over who is the inventor of an invention involving AI activity.

## 2 Promoting R&D and other collaboration between publishers, AI-focused SMEs and academia:

- Many publishers (particularly SMEs) do not have the capacity, resources or skills to experiment with and develop AI technologies in-house, and those with some capacity still value partnership or collaborative working. Collaborative working between business and academia is at the heart of the UK Government's Industrial Strategy.<sup>1</sup> The Strategy includes specific policies focused on the Creative Industries,<sup>2</sup> and on promoting AI-led innovation in the UK.<sup>3</sup> At present, though, despite our evidence on the potential of AI in a key part of the creative economy, publishing's voice appears to be relatively unrepresented in this critical policy debate. We recommend that the **publishing sector engages strongly** with the developing Industrial Strategy, ensuring that the industry is fully represented in future policy discussions which will enable publishing's needs to be better addressed with tailored support. **Government should also actively engage directly with the sector** – for example, encouraging participation in bidding for collaborative funding through Challenge Funds working with academics and tech-focused SMEs to drive and enable future AI-led innovation.

## 3 Helping publishing SMEs access AI investment finance and skills:

- Our industry survey suggests that SME publishers (particularly micro-enterprise) recognise the potential benefits of AI investment but lack the resources they need to invest. The costs of researching, acquiring and implementing AI solutions and AI skills were identified as significant investment barriers by SME publishers that responded. We recommend **further engagement between the Publishers Association and its SME members**, for example through a taskforce focused on addressing SME AI investment issues. Evidence from this should inform future engagement with government exploring how to improve SME's ability to invest in and benefit from cutting edge technologies such as AI.

<sup>1</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf)

<sup>2</sup> <https://www.thecreativeindustries.co.uk/media/462717/creative-industries-sector-deal-print.pdf>

<sup>3</sup> <https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/industrial-strategy-the-grand-challenges#artificial-intelligence-and-data>

# 1

## Introduction

This report for the Publishers Association (PA) represents, to our knowledge, the first piece of systematic analysis seeking to document the role of Artificial Intelligence (AI) and Machine Learning (ML) for UK publishing. It seeks to demonstrate the ways in which the UK publishing industry is innovative, dynamic and working to harness advances in AI and ML technology to deliver benefits to the sector, its consumers and wider society.

The scope of our research is broad and forward looking, covering these areas:

- Defining Artificial Intelligence and Machine Learning in the context of the publishing sector and exploring sub-sector and regional differences;
- Developing a taxonomy that describes where investment in AI is occurring and the benefits to publishers and the wider economy;
- Identifying the investment barriers being faced by publishers and issues that are arising as a result of AI investment;
- Assessing the impact that AI may have on publishing in the future; and
- Making recommendations to help ensure that the sector can make best use of AI and ML in driving future innovation and productivity growth.

Driving creative industry performance and investment in Artificial Intelligence are high priorities for UK government policy.<sup>4</sup> We hope that this report will provide timely evidence for industry stakeholders and policy makers to help inform future decision-making around the creative industries and employment of AI technologies.

---

<sup>4</sup> As evidenced by the government's Industrial Strategy ([https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf)) and specific interventions relating to this strategy, including Sector Deals both for creative industries (<https://www.gov.uk/government/publications/creative-industries-sector-deal>) and AI (<https://www.gov.uk/government/publications/artificial-intelligence-sector-deal>).

## 1.1 How we define publishing for this study

We define the publishing sector as all companies conducting academic, education and consumer publishing activities. We exclude newspaper, catalogue and magazine publishing and also software and computer games. This definition reflects the breadth of organisations supported by the Publishers Association and aligns with previous analysis we have carried out (Figure 1).



Figure 1.

Source: Taken from Frontier Economics (2017), “The contribution of the publishing industry to the UK economy”, 2017

Note: Areas of boxes are proportional to GVA. Books and academic journals GVA estimating using the Publishing Yearbook turnover and ABS GVA/turnover ratio. Other GVA figures directly from ABS. All figures from 2015. Although the PA definition includes the publishing of translated books, it excludes translation activities when they are outsourced to translation agencies.

Previous analysis by Frontier Economics for the PA demonstrated the important economic contribution made by this part of the UK’s publishing industry – directly contributing £3.2 billion of GVA in 2016 and supporting 29,000 jobs with high average productivity.<sup>5</sup> In 2019, the value of UK publishers’ sale of books, journals and rights was £6.3 billion, an increase of 20% compared with sales in 2015, with exports representing 59% of sales<sup>6</sup>. The 2017 Bazalgette Review<sup>7</sup> also remarked on the increasing importance of the creative industries to the UK economy due to its characteristics of high economic growth, high levels of investment in innovation and technology, a labour market that is relatively immune to the threats from automation and the value it generates from exports.

<sup>5</sup> Frontier Economics (2017), “The contribution of the publishing industry to the UK economy” (<https://www.publishers.org.uk/wp-content/uploads/2020/02/Contribution-of-the-Publishing-Industry-to-the-UK-Economy-2017-1.pdf>)

<sup>6</sup> The Publishers Association Yearbook 2018.

<sup>7</sup> <https://www.gov.uk/government/publications/independent-review-of-the-creative-industries>

Publishing also supports a wider value chain – from blue skies research and authors (who play a dual role of supplier and customer) through to warehousing and printing service providers, marketing, retail and education. This report explicitly considers the implications of the use of AI and ML by UK publishers for this wider set of supporting industries. We also consider how the quality and reach of high value creative inputs from publishing could be affected by the use of AI and ML. As shown in previous research, these creative inputs from publishing represent significant benefits for the UK's wider creative industries.<sup>8</sup>

## 1.2 Current evidence on the use of AI in the UK creative industries

One of the aims of this study was to compare the use of AI in publishing with other creative industries. Our review of the evidence, however, suggests that this is not a well-researched subject – we found no industry consultations or surveys to shed systematic light on this issue. As a result we are not able to use the findings from this study to make general comparisons between publishing and other creative industries.

The two studies we identified were analyses of surveys focused on the Design and Advertising industries. Both were global (not UK-specific) surveys of industry professionals.

- In **Advertising** key results suggested that around one in eight businesses were already using AI, increasing to more than a quarter of 'top performing' organisations. Around one-third of businesses were planning to increase their use of AI in the future. Key barriers to using AI included a 'lack of perceived need', 'knowledge of how to use AI', and 'lack of resources'.<sup>9</sup>
- In **Design** key results suggested that just over a third of businesses felt that AI and ML will be extremely important for creative professionals and a further quarter felt that AI and ML will be quite important. However there was no specific analysis of the use of AI or ML in the sector.<sup>10</sup>

---

<sup>8</sup> See Frontier Economics (2018), "Publishing's Contribution to the Wider creative industries" (<https://www.frontier-economics.com/media/2503/publishings-contribution-to-the-wider-creative-industries.pdf>). The official DCMS definition of the creative industries includes "those industries which have their origin in individual creativity, skill and talent and which have a potential for wealth and job creation through the generation and exploitation of intellectual property". At a sector level, creative industries are Advertising and Marketing, Architecture, Crafts, Product Design, Graphic Design and Fashion Design, Film, TV, Video, Radio and Photography, IT, Software, Video Games and Computer Services, Publishing and Translation, Museums, Galleries and Libraries, Music, Performing Arts, Visual Arts and Cultural Education.

<sup>9</sup> Sourced from the Digital Trends 2018 survey of 13,000 Advertising professionals in Europe, North America and Asia Pacific - <https://www.adobe.com/content/dam/acom/en/modal-offers/pdfs/0060629.en.aec.whitepaper.econsultancy-2018-digital-trends-US.pdf>

<sup>10</sup> Sourced from Creativity and Technology in the Age of AI, drawing on evidence from 110 interviewees in Europe, US, Japan - <https://www.pfeifferreport.com/essays/creativity-and-technology-in-the-age-of-ai/>



## 1.3 Defining AI in the context of publishing

There is no single, standard definition of AI. This stems from the fact that the technologies it embraces are constantly changing, and also that the way AI is described depends on its application, with experts in different disciplines bringing their own perspective and terminology.<sup>11</sup>

Our understanding is that AI is a broad term that covers a range of applications whose purpose is to enhance existing intelligence. The goal of AI is not to replace human intelligence; instead it adds value through its superior processing power, ability to memorise and calculate. Freeing up humans from the need to (and costs of) carrying out repetitious tasks should, in theory, spur creativity and improve productivity.

In the context of publishing, AI covers Machine Learning (ML) technologies which are combined with text, image and voice related technologies such as Natural Language Processing (NLP), Voice Recognition or Computer Vision. This combination of technologies can be deployed to help search through and join up content more effectively, for example based on a set of key words. More recently, Deep Learning (a sub-category of Machine Learning which seeks to replicate the neural networks of the human brain) is being applied to large amounts of complex data (for example, academic research papers, student learning outcomes or fiction novels) to spot patterns, generate new hypotheses, learn how to summarise text, and create novel content.

In consultation with publishing stakeholders,<sup>12</sup> we have developed the following working definition of AI for this study, recognising that what AI includes is constantly changing and can be interpreted differently depending on the specific application.

## DEFINING ARTIFICIAL INTELLIGENCE FOR THE PUBLISHING SECTOR

Artificial Intelligence is a broad term that covers multiple technologies which enable computers to sense, comprehend, act and learn. All current uses of AI in the publishing industry involve Machine or Deep Learning - either alone, or in combination with other technologies such as Natural Language Processing, Voice Recognition or Computer Vision.

<sup>11</sup> Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). *Intelligence Unleashed. An argument for AI in Education*. London: Pearson.

<sup>12</sup> We developed this definition through conversations with publishers based on their in-house working definitions. We tested the definition out as part of the industry-wide survey. 96% respondents agreed that AI is a broad term covering multiple technologies that enable computers to sense, comprehend act and learn. 88% respondents agreed that all applications of AI in publishing used ML or DL alone or in combination with NLP, CV or VR. Publishers that did not agree indicated that they did not have a working definition or that they used a Machine Learning-based definition. Our broad definition of Machine Learning (of which DL is a subset) addresses this issue.

## 1.4 Our methodology and evidence base

We adopted a four-step approach to this study, summarised below.



Figure 2. Source: Frontier Economics

This study is informed by significant amounts of primary and secondary evidence collected over a period of four months:

- an extensive **desk-based literature review**;
- a total of **7 scoping interviews with publishing stakeholders** covering a range of publishing sub-sectors;
- an **online industry survey** that was sent to all UK Publishers Association member organisations; and
- a total of **3 in-depth case studies** involving both interviews and desk research to understand particular current applications of AI in publishing.

Our study also benefited from significant constructive feedback from stakeholders during the development of the taxonomy and research materials. We extend our thanks to all those who participated in this research.

### 1.4.1 The literature review

We identified relevant papers, book chapters, articles and industry surveys at the outset of the project. A full bibliography of the review can be found in Annex C. These helped us to understand how publishers are applying AI, whether there are geographical or sub-sectoral differences in the nature and intensity of AI investment and expectations as to future applications of AI.

This evidence provided a foundation for developing a **taxonomy** of how publishing is being or could be used in publishing. The evidence from the review helped to identify and categorise the nature, application and benefits of AI technologies in the different areas of publishing. We also identified and reviewed material relating to the ongoing World Intellectual Property Office (WIPO) consultations with regard to the implications of AI for Intellectual Property (IP) and are aware that the UK has commenced its own consultation process on this subject.

## 1.4.2 Stakeholder scoping interviews

We provided early drafts of our taxonomy to seven publisher stakeholders for feedback which was gathered using individual semi-structured telephone interviews lasting between 30 minutes and one hour. These interviews explored how the taxonomy applies to their business operations and investment plans for AI. We also discussed how publishers define AI, the investment barriers they face and their views on the role of policy in addressing these. The evidence from these conversations was used to finalise the taxonomy. It also fed into the design of the survey and case studies and helped to inform overall recommendations.

Our interview guide for these conversations is provided in Annex A.

## 1.4.3 The industry survey

We are not aware of up-to-date, comprehensive, industry-wide primary evidence on the role of AI in publishing, or in the wider creative industries for the UK. To address this we designed and implemented an industry survey to address this gap. The survey was tailored on the basis of the finalised taxonomy and the wider research questions this study was seeking to address.

The survey was distributed online to all 135 of PA's members<sup>13</sup> on 17 February 2020. Fieldwork ran for one month (closing on 16 March 2020). Overall, 31 publishers responded (a 23% response rate) and 24 publishers completed every question relevant to them (18% of members).

We received a significant response from large publishers (12 complete responses from those employing 250 or more full-time equivalent (FTE) staff worldwide), representing approximately two-thirds of PA's large members<sup>14</sup>. Recent national statistics suggest that there are, at most, only around 50 to 60 UK owned large publishing organisations in the UK<sup>15</sup> (employing 250 or more FTE staff in the UK). Whilst these statistics are based on slightly different definitions of large publishing companies, it suggests that our results are likely to provide a good representation of this particular part of the industry.<sup>16</sup>

---

<sup>13</sup> PA had 135 members as of 31 March 2020.

<sup>14</sup> This is an approximate proportion. The Publishers Association uses their members' annual UK turnover to classify them by size. As of 31st March 2020 18 members had annual UK turnover greater than £25m.

<sup>15</sup> Note: This is only an approximate comparison. Our definition of large publishing organisations for the survey was based on annual global turnover and therefore may include organisations that would not be considered as large by UK national statistics. Also, the ONS definition of publishing is broader than our working definition (see footnote below).

<sup>16</sup> Source: ONS Business Population Estimates for the UK and regions 2019. Table 7, shows there to be 55 large organisations conducting publishing of books, periodicals and other publishing activities (a wider scope of publishing industries than considered by this study) [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/836562/BPE\\_\\_2019\\_detailed\\_tables.xls](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/836562/BPE__2019_detailed_tables.xls)

Nearly all of the remainder of responses came from micro-sized publishers (12 partial responses, 10 full responses). We also received 3 full responses from small publishers (employing 10-49 FTE staff globally) and 1 full response from a medium-sized publisher (employing 50-249 FTE staff globally). Our survey evidence is therefore not representative of the UK's small to medium sized publishers.

Some of the survey questions focused on publishers who were 'using or researching any form of AI'. Just 13 publishers stated they were researching or using AI and so the sample for this part of our analysis is much smaller. 12 of these 13 publishers are large and so we have only been able to draw conclusions here for large publisher AI activity (and note that this also suggests that micro enterprise publishers are not yet investing). More research is required to understand the AI research and investment behaviour of SME publishers.

We received a broad response in terms of different publishing sub-sectors including academic, consumer and education; and domestic and international publishers. Annex B provides further detail on the characteristics of the survey and respondents.

## 1.4.4 The case studies

We constructed a series of deep-dive case studies to illustrate the diverse range of cutting-edge applications of AI at different stages of the value chain and for different types of publisher:

- **Case study 1: AI and academic content generation** – illustrated using the example of Springer Nature's Lithium Ion Batteries publication.
- **Case study 2: AI and publisher content recommendation** – illustrated using the example of Elsevier's Mendeley Suggest and Taylor & Francis' recommendation platforms.
- **Case study 3: AI and education publishing product development** – illustrated using the example of McGraw Hill's ALEKS and SmartBook platforms.

Each of these case studies draws on primary evidence from stakeholder interviews and desk research (including materials provided by the case study publishers).

## 1.5 Report structure

The remainder of this report is structured as follows:

- Section 2 describes our **taxonomy** of AI in publishing – this provides a framework for the remainder of our analysis. It draws on our literature review and stakeholder evidence.
- Section 3 describes our evidence on **how UK publishers are using AI**. It draws on the primary evidence we collected from an online industry survey, the stakeholder interviews and case studies.
- Section 4 considers the **AI investment barriers** that publishers are facing. This also draws evidence from the online industry survey, stakeholder interviews and case studies.
- Section 5 is forward looking, presenting evidence on publishers' perceptions on the **future impact of AI investment** for the industry.
- Section 6 sets out **our recommendations** for supporting continued AI investment.



# 2



## A taxonomy of AI in publishing

To better understand the extent to which publishers are investing in AI we designed a generic framework which conceptualises how AI is used and the benefits it can bring. Our aim was to produce something simple, which can be quickly understood and adapted to fit the different sub-sectors (education, academic and consumer) publishing.

As discussed in Section 1.4, the taxonomy was developed based on an extensive literature review and consultation with publishing stakeholders.

Our analysis and interviews did not suggest there were any particular **geographic** variations in AI investments or its potential impact on the industry. This includes variations between the UK and other countries, or variations in different regions of the UK. Our taxonomy is therefore likely to have broad applicability to the role of AI in global publishing.

### KEY FINDINGS

Publishers tend to adopt a **benefit-led approach** to their AI investment decisions. Other factors relevant to the decision making process include the nature of the technology and the value chain stage the investment is associated with. Our 'value chain – technology – benefits' framework reflects this and describes AI technologies in terms of their position in the value chain, the nature of the technology and the likely investment benefits.

Many of the **AI technologies** and **benefits** we identified appear to be **common to all publishing sub-sectors**. The technology focus is on combining text and image recognition technologies such as NLP with ML. Benefits are often those associated with task automation and also the ability to rapidly and accurately process large volume of data (cost and time savings, strategic insight, risk reduction, for example) and also wider benefits such as increased innovation activity in other sectors.

The range of current AI uses cases by the industry is broad. Some of the most **common industry-wide applications** include identifying market trends to inform content acquisition, carrying out copyright infringement checks, running language and grammar checks, recommendation engines and demand forecasting to inform marketing strategy and manage stock levels. Some academic publishers also use AI to help researchers organise and share their content and identify relevant research. A few education publishers are creating interactive adaptive learning solutions and there is evidence of ML being used to automate consumer book pricing.

## 2.1 Taxonomy structure

Our AI in publishing taxonomy is structured around a generic publishing value chain. As Figure 3 illustrates, it provides a high-level representation of:

- the different sub-classes of **AI technology** that are being applied by publishers at each stage of the **value chain**; and
- the nature of **benefits** that publishers expect to gain from their investment as well as a broad assessment of the wider society level benefits that may follow on a direct result.

Figure 3.

Source: Frontier Economics



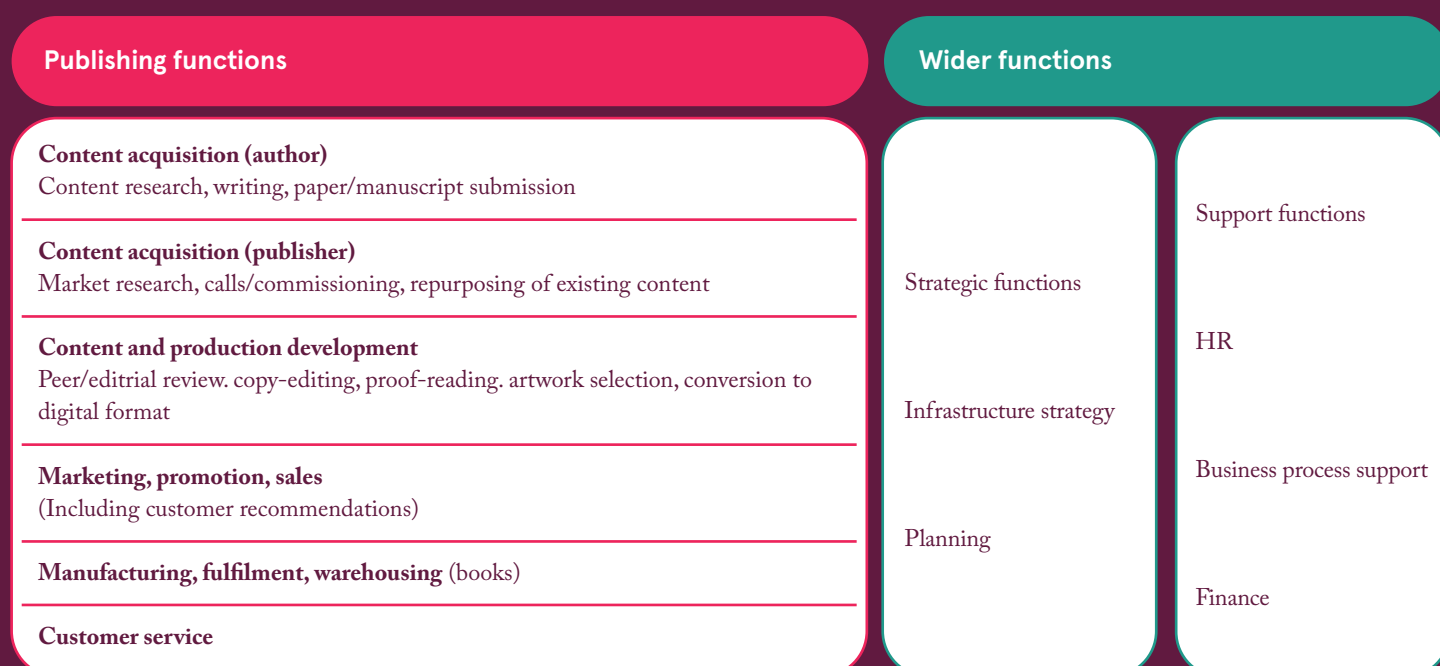
Publishers told us that this approach resonated with the way they think about their AI investment activity. In some cases, we heard that the investment decision-making process tends to be benefit-led rather than technology-led (“Can we address this issue, be more effective in this area or deliver significant benefits to our customers by employing AI?”, rather than “What can we achieve with this AI technology?”). Technology-led investment decisions tended to be driven by a need to establish ‘technology fit’ within the existing organisational structure, data availability, cost and by technology readiness.

### 2.1.1 Value chain activity

Figure 4.

Source: Frontier Economics

Figure 4 depicts the generic value chain. Publishing-specific functions are shown on the left hand-side and wider, non-publishing specific functions on the right.



The value chain describes the publishing production process starting with authors or researchers who **acquire** the content (for example, published research materials) they need to write research articles, chapters or books intended for publication. Publishers **acquire** their content through calls or commissioning (in some cases this is supported by market research). Existing content can also be repurposed<sup>17</sup>. Content is **developed** into a marketable product through an editorial process and type-setting, artwork, format and language translation activities. **Marketing** can occur through a number of channels (e.g., website and recommendation engines), drawing on market intelligence to inform the strategy. The final product may be **manufactured**, warehoused and delivered to retailers (we understand many publishers outsource this function).

The value chain in Figure 4 also depicts publishers' wider functions of **customer service**, **strategic functions** such as strategy and infrastructure planning and **support functions** such as HR and finance. Although not the core focus, these functions represent an important element of publishers' businesses and are within the scope of this study.

## 2.1.2 AI technology

Figure 5 briefly describes the AI technologies that are relevant to the taxonomy.

AI technology	Description
<b>Machine Learning (ML)</b>	Technology that enables computer to be programmed so that they can learn, unaided, from data to make better predictions. Supervised ML requires classified and labelled data whereas unsupervised ML can learn from unlabelled data
<b>Deep Learning (DL)</b>	A form of ML employing neural networks (interconnected data sets) intended to replicate the human brain capable of analysing large volumes of data to recognise patterns and make decisions
<b>Natural Language Processing (NLP)*</b>	Converts text into data (or vice versa) enabling computers to identify words and combinations of words and to create written output.
<b>Text and Data Mining (TDM)</b>	Technology that enables the discovery of knowledge from unstructured data or text.
<b>Image recognition*</b>	The ability to recognise an object or figure in a digital image or video.
<b>Speech recognition*</b>	The ability to recognise and transcribe the human voice into data for analysis
<b>Decision management</b>	Machines that apply rules and logic to AI systems enabling automated decision making
<b>AI optimised hardware</b>	Using AI to automate physical tasks (such as using robotics in stockrooms)

Figure 5. Source: Frontier Economics

Note: \* Some cognitive technologies listed here such as NLP can be employed with and without AI technologies. In this case we are referring only to instances where these technologies are employed in combination with AI technologies

Our discussions with publishers revealed that a similar set of technology types apply to all publishing sub-sectors. The technology focus is on combining text and image recognition technologies with AI technologies (a mix of machine learning and deep learning).

<sup>17</sup> We understand this is common practice in education publishing.

## 2.1.3 AI benefits

Overall, we found that a similar set of benefit-types apply to all publishing sub-sectors.

Many of the benefits that publishers described to us were those associated with task automation. AI's ability to process large volumes of (text and numeric) data within significantly shorter time frames, and more accurately than is possible for humans, can generate significant time-saving benefits. Because more data can be processed and scope for human error is removed, AI technologies can also improve the analytical quality of an output, generate additional strategic insight and in some cases can be used to check legal compliance to reduce risk exposure.

Task automation reduces the labour effort and labour costs associated with a function. We asked publishers whether they were concerned about AI investment resulting in a skills shift within the industry. The majority of publishers currently do not appear to be greatly concerned<sup>18</sup>. Outside of the publishing industry, our research provided examples of AI-driven automation within publishing enabling individuals to switch to higher value tasks within the same role. One key example is in teaching, where AI can carry out routine tasks freeing up teachers to spend more time teaching<sup>19</sup>.

These function-specific internal benefits lead to wider, second order benefits, including increased innovation levels (leading to higher quality products or services) and increased sales volumes. UK publishing companies may also become more efficient, driving UK economic growth. AI also has the potential to drive significant customer benefits by, for example, saving consumers, researchers and teachers time, delivering them with more accurate and up to date information and providing them with more choice.

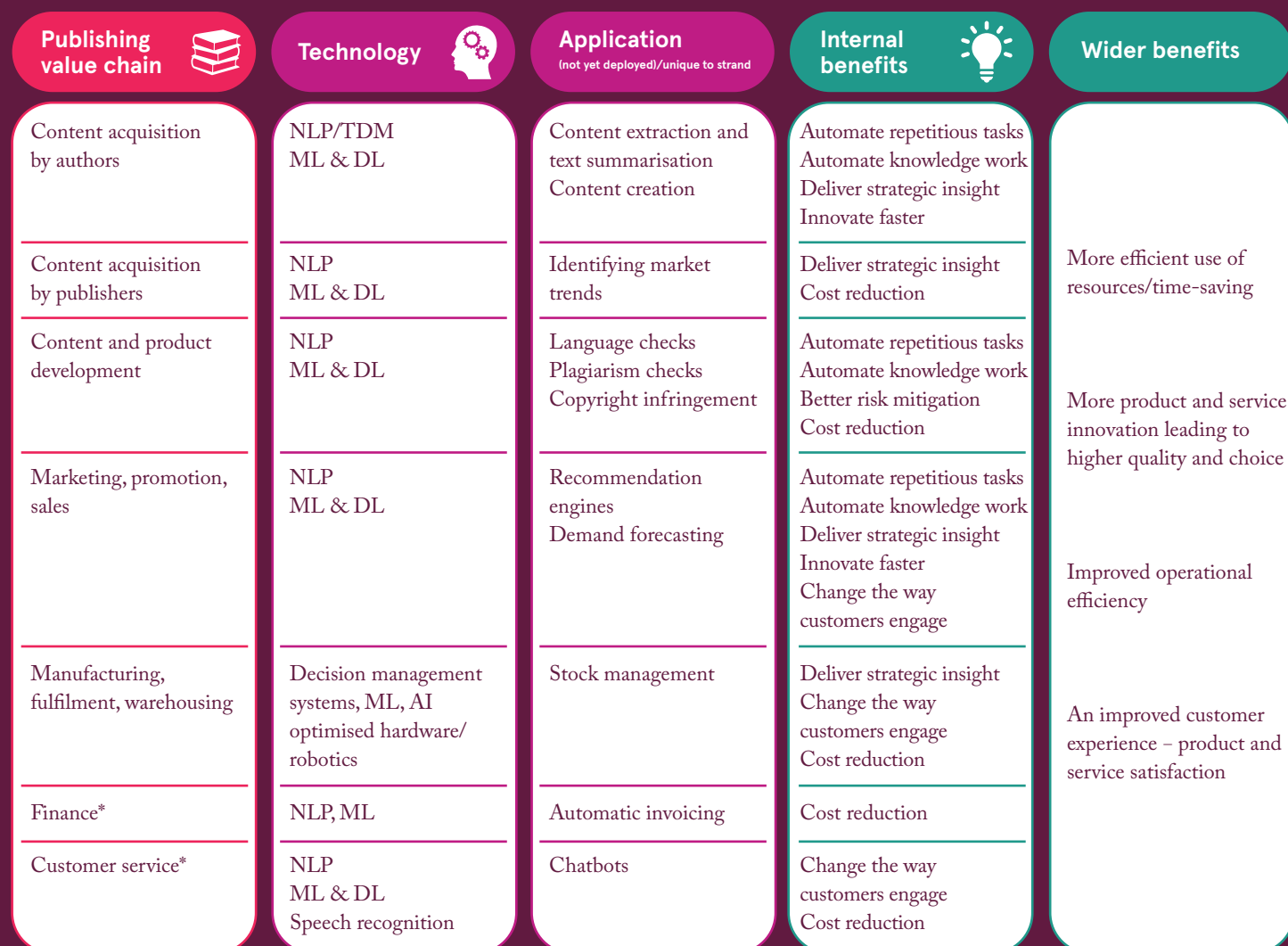
---

<sup>18</sup> 3 out of 26 publishers that responded to the question in our industry survey: 'To what extent are you concerned about a skills shift' stated 'greatly concerned', a further 10 stated 'slightly concerned'.

<sup>19</sup> Luckin, R., Holmes, W. 2017. Intelligence Unleashed: An argument for AI in Education. <https://discovery.ucl.ac.uk/id/eprint/1475756/>

## 2.2 The overall taxonomy

The role of AI in publishing be represented to a large extent by a broad taxonomy for academic, education and consumer publishing. This is shown in Figure 6.



**Key:** NLP – Natural Language Processing TDM – Text and Data Mining ML – Machine Learning DL – Deep Learning

**Figure 6.**

**Source:** Frontier Economics

**Note:** Stakeholder evidence suggests that manufacturing, fulfilment and warehousing is mostly relevant to education and consumer manufacturers.

\* refers to non-publishing-specific value chain activities

### 2.2.1 Describing the role of AI at each value chain stage

We describe the technologies, applications and benefits for each value chain stage in more detail. After this we identify sub-sector-specific variations.

#### AI in author content acquisition

**Text mining techniques** are relatively common in publishing. These employ NLP to recognise key words or strings of words to identify relevant content for authors (including researchers). These techniques deliver benefits by **automating repetitious**



**tasks and knowledge work** (such as searching through large volumes of websites or journal articles for the same information), and can significantly reduce the time that authors spend capturing underlying data and resources to inform the writing process. As a result authors can devote more time to gain strategic insight, create or innovate better and faster.

Other less mature examples include **text summarisation** and **content creation** technologies. Both build on the content extraction process to convert the machine-extracted material into readable text. This application is being developed by all publishers but the most mature examples are recent and tend to be from academic publishers (for example, Springer Nature's Lithium-Ion Batteries publication<sup>20</sup> and Kogan Page's book 'Superhuman Innovation')<sup>21</sup>. In consumer publishing, the prospect of releasing AI-generated books was described as 'very likely' by one stakeholder but until more is known about the underlying technology<sup>22</sup>, the precise timing of this is impossible to predict, and is unlikely to be within the next five years.

### AI in publisher content acquisition

Publishers can use a combination of NLP and ML to search through published online content to identify clusters of key words (topic areas) and **identify market trends**. In some cases publishers will use this technology to search through their existing content to spot where there are gaps in their current offer. In both cases this information can deliver **strategic insight** that informs their commissioning strategy enabling them to deliver content that better meets customer needs. Automating the search process in this way also **reduces publishers' costs**.

### AI in content and product development

A combination of NLP and/or image recognition technology with ML technology can be used by publishers to flag papers with similar sounding paragraphs and sentences to **identify plagiarised content** and to carry out **copyright infringement checks** by scanning materials published in the public domain. Several stakeholders told us that they are already doing this. The key benefits associated with this are an increase in the quality of published material available to consumers, risk mitigation and an increase in publishers' revenues.

Most of the publishers are already using ML to run **language (spelling and grammar) checks**. ML can also help to **screen and edit manuscripts**. In academic publishing it can comprehend the unique properties of scholarly content and incorporate the requirements of specific publications by either making direct changes or signalling to editors the parts that need to be reviewed. Examples from our stakeholder interviews include running automatic English quality assessments on content, checking image labelling for accuracy and predicting whether a paper needs to be copy-edited. There are significant potential benefits from this application of AI, particularly the **automation of knowledge work** which should in turn drive cost reductions for publishers and particularly for consumer publishing, using AI to **translate** content for other markets.

<sup>20</sup> Beta Writer, 2019. Lithium-Ion Batteries: A machine-generated summary of current research.

<sup>21</sup> Duffey, C. 2019. Superhuman Innovation – Transforming business with artificial intelligence. Kogan Page. EAN: 9780749483838

<sup>22</sup> One stakeholder mentioned Open-AI's 'GPT-2' output detection model. This large language model was released in stages throughout 2019, and is considered capable of generating realistic paragraphs of text. See for example, <https://medium.com/syncedreview/openai-releases-1-5-billion-parameter-gpt-2-model-c34e97da56c0>

### AI in marketing promotion and sales

Recommendation engines that use NLP and ML are being employed by publishers to help their readers identify the content most relevant to their needs. These platforms analyse data on readers' explicit stated preferences, their research behaviour and those of others with similar profiles. For academic researchers this can lead to benefits such as **improved strategic insight**, leading to the creation of **better quality research** themselves. In consumer publishing, these platforms have the potential to **reduce consumer search costs**, generate time-saving benefits and potentially higher purchase satisfaction whilst **increasing publishers' sales volumes**. The education stakeholders we spoke to described recommendation technology in a slightly different context (see the discussion on product development in Section 2.2.2).

Some stakeholders also told us they were using, or researching, the potential of AI to automate their knowledge work in terms of **demand level forecasting**, using ML to analyse their historic marketing data. This has the potential to **inform publishers marketing strategies** and also feeds into stock management practices (discussed next).

### AI in manufacturing, warehousing and fulfilment

Linked with **demand forecasting** above, the literature evidence and our stakeholder interviews provided examples of publishers using AI to predict demand and improve **stock management** - one of the biggest cost items for consumer publishers.<sup>23</sup>

Our stakeholder interviews did not provide evidence of publishers using AI to **optimise hardware or robotics** for manufacturing and warehousing, but there is evidence of AI starting to be used by organisations such as Amazon to reduce their logistics costs, although full AI-driven automation is considered to be several years away.<sup>24</sup>

## 2.2.2 Sub-sector-specific technologies and benefits

Our literature review and stakeholder interviews revealed some AI in publishing use-cases that appear currently to be specific to a particular sub-sector. As AI technology develops, publishers in the other sub-sectors may adopt these solutions. With this caveat in mind we present our sub-sector taxonomies with a brief description of our evidence on sub-sector specific AI applications.

---

<sup>23</sup> Clark, G. and Phillips, A. 2019. Inside Book Publishing. Routledge.

<sup>24</sup> <https://www.theverge.com/2019/5/1/18526092/amazon-warehouse-robotics-automation-ai-10-years-away>

## Academic publishing

Figure 7 presents the academic publishing AI taxonomy.



**Key:** NLP – Natural Language Processing TDM – Text and Data Mining ML – Machine Learning DL – Deep Learning

Figure 7.

Source: Frontier Economics

Note: Red text indicates applications or benefits that are specific to academic publishing.

The peer review process straddles content acquisition and development. We place it in development because of the significant editorial role of peer review (which fits alongside copy editing and proof reading).

Applications of AI that are specific to academic publishing include:

- **Reference management and sharing:** At the top of the value chain, AI enhanced tools enable researchers to download, organise and share their downloaded content with their wider research community. Benefits of this technology include facilitating researcher collaboration, knowledge management (making the overall research process more effective) and enabling knowledge discovery.<sup>25</sup>

<sup>25</sup> <https://blog.mendeley.com/category/mendeley-use-case/>

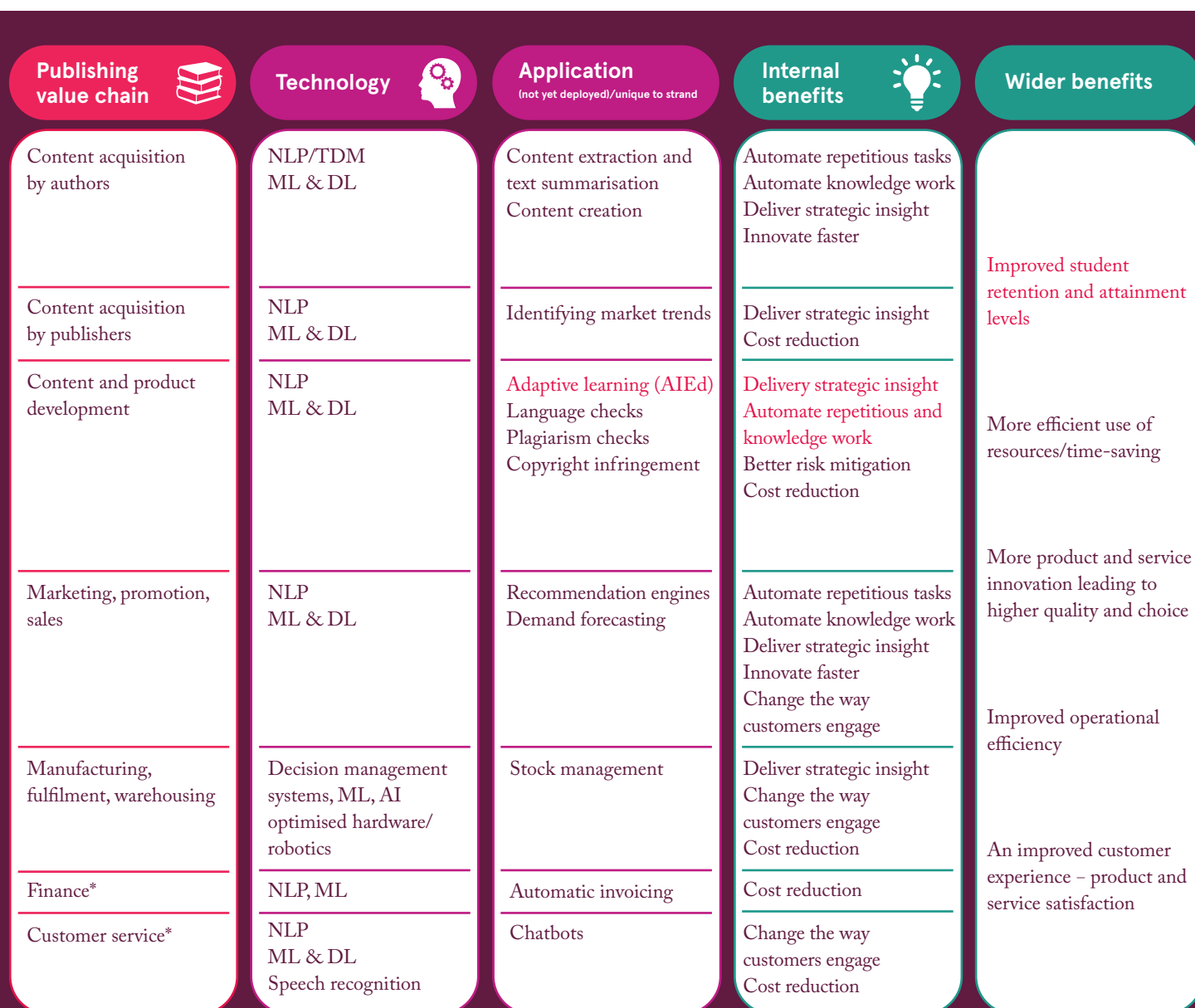
- **Assisting with clinical decision making and generating research hypotheses:** In the next value chain stage our stakeholder interviews provided examples of using AI to create tools to assist clinical decision making and to generate new hypotheses for medical and scientific research. As well as NLP and ML, computer image recognition technologies are required. This technology could deliver significant strategic insight for research, reducing costs and delivering welfare benefits to society through, for example, helping to discover new medicines.
- **Classification and indexing:** Many academic publishers are using metadata tagging to enable their content to be discoverable by search engines and browsers. Metadata tagging creates a term that describes a keyword or phrase and assigns these ‘tags’ to the document’s digital assets. The tags don’t appear to the user, but are in the source code. It enables their content to be easily identified and indexed, saving time and removing human bias (by replacing current manual, subjective tagging methods), facilitating knowledge sharing and creation and driving publishers’ sales.
- **Identifying relevant peer reviewers:** Stakeholders described peer review as a time and resource intensive process, with each paper usually requiring at least three peer reviewers). As well as saving time, employing a combination of NLP and ML to identify academic researchers with the required expertise, can enable publishers to widen their search and find new peer reviewers<sup>26</sup> potentially increasing the quality of review. The literature suggests that AI itself could also be used to peer review academic articles. Some publishers we spoke to raised concerns about using AI in this way. The key concern is creating a potential peer review ‘black box’ whereby other issues such as programmed-in bias, or inaccurate reporting may be hard to detect.<sup>27</sup>
- **Verifying research results:** AI can be used to verify research results and detect whether data has been modified or generated to fabricate an outcome. This could increase the quality of academic research that builds on this evidence, potentially generating cost and welfare benefits depending upon the outcome of subsequent research.

<sup>26</sup> Source: BioMed and Digital Science. 2017. What might peer review look like in 2030? [https://figshare.com/articles/What\\_might\\_peer\\_review\\_look\\_like\\_in\\_2030\\_/4884878](https://figshare.com/articles/What_might_peer_review_look_like_in_2030_/4884878)

<sup>27</sup> For example, University of Trieste researchers carried out a study in which they generated fake peer reviews and asked academics to say whether they agree with them. In a quarter of the cases academics agreed with fabricated. <https://medium.com/digital-publishing-strategy/the-impact-of-ai-on-peer-review-628f1318b18c>

## Education publishing

Figure 8 presents the education publishing AI taxonomy.



**Key:** NLP – Natural Language Processing TDM – Text and Data Mining ML – Machine Learning DL – Deep Learning

**Figure 8.**

Source: Frontier Economics

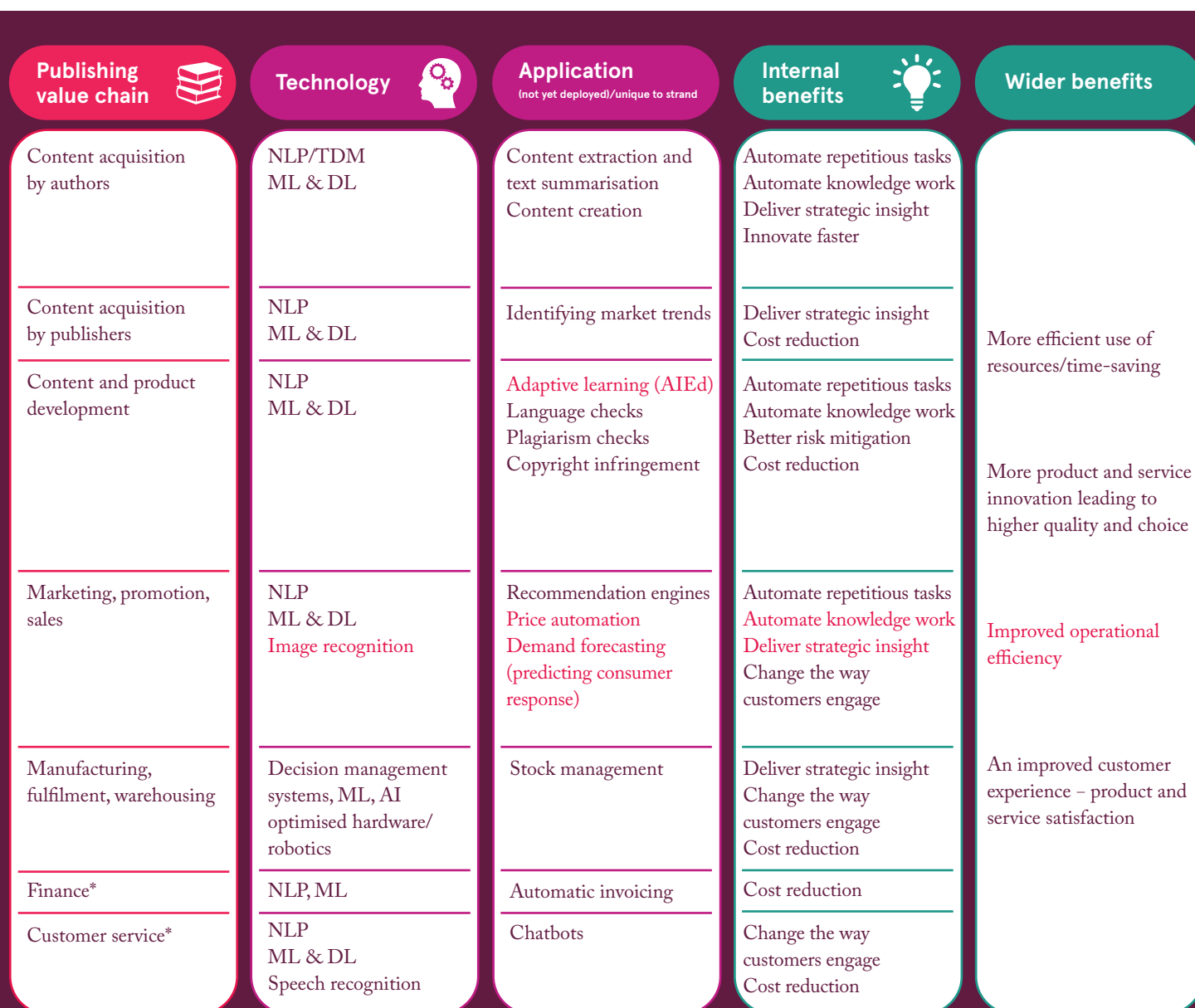
Note: Red text indicates applications or benefits that are specific to education publishing.

Our evidence on how AI is being applied by education publishers highlights the importance of product development and design for education publishers, and identifies an education specific application of AI to create **adaptive learning solutions** (commonly referred to as AIEd<sup>28</sup>). AIEd links publishers' student performance data to educational content. There is significant innovation in this area and considerable interaction between education publishers and innovative Ed Tech and AI tech focused research organisations. A number of education publishers have invested in this area and expect the technology to drive benefits such as improved student retention and attainment levels, and improved use of classroom time and overall time saving for educators.

<sup>28</sup> For a broader discussion of AI in education see Luckin, R., Holmes, W., Griffiths, M. & Forcier, L. B. (2016). Intelligence Unleashed. An argument for AI in Education. <https://static.googleusercontent.com/media/edu.google.com/en//pdfs/Intelligence-Unleashed-Publication.pdf>

## Consumer publishing

Figure 9 presents the consumer publishing AI taxonomy.



**Key:** NLP – Natural Language Processing TDM – Text and Data Mining ML – Machine Learning DL – Deep Learning

**Figure 9.**

Source: Frontier Economics

Note: Red text indicates applications or benefits that are specific to consumer publishing.

Our consumer-publishing specific evidence highlights the importance of AI and consumer publishers' marketing, promotion and sales activities. Some publishers are using ML to **automate the price** they set to retailers based on the performance of a title and data on historical sales of comparable titles<sup>29</sup>. One publisher also told us they were applying AI to **image recognition** technologies to predict consumer response to book cover designs or marketing materials.

<sup>29</sup> Bookwire (a European e-book and audiobook distributor) is using ML to recommend ideal prices and to inform their book promotion strategy, Gould Finch (2019) suggest this has boosted publishers' backlist sales by up to 20%.



## 2.2.3 AI technologies and their benefits outside of core publishing industry activities

AI can be applied to non-publishing-specific tasks such as **finance** to automatically raise invoices. The stakeholder interviews reveals that some publishers are using AI in this way to generate significant cost reductions. **Chatbots** use a mix of NLP, speech recognition and text generation algorithms and can be placed on publishers' websites to assist with **customer service**. In consumer publishing they serve a dual purpose: helping readers identify books and collecting data for trend analysis. One example of this is Penguin Random House's Reader Rewards Loyalty Programme, launched in 2019<sup>30</sup>.

---

<sup>30</sup> <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/79824-prh-creates-reader-loyalty-program.html>

# 3



## Evidence on the role of AI in publishing

Our taxonomy provides a framework to assess the role of AI in publishing in the UK, both in terms of the current position and the potential future position.

In this Section, we assess:

- Publishers' current and planned AI investment levels (for all sized publishers); and
- Where investment is occurring within the publishing value chain and the benefits this is generating (here we focus on large, investing publishers).

### KEY FINDINGS

Our evidence on investment activity to date comes from a group of 12 large UK publishers (accounting for around two-thirds of the PA's large members). We refer to this group as 'large AI-active publishers'. A large part of our analysis on publishers' AI investment activities focusses on this group. We did not gather sufficient evidence from SME publishers to draw firm conclusions on AI investment for this group.

Our evidence from all publishers suggests that **AI investment levels are increasing year-on-year** and that large publishers are leading investment activity. The majority of (all) publishers started to invest in AI in 2017 and now have small teams of around 1-5 FTE staff (around half of whom are based in the UK) focused on AI research.

Focusing on investment by large AI-active publishers we find that:

- **AI is being applied to all parts of the value chain except customer service.** Whilst the highest proportion of these publishers are already using AI (some extent) for content acquisition and development, three-quarters are using or plan to introduce marketing and sales solutions within the next two years.
- **The most common AI-assisted task to date is content classification** (for example, using meta data tagging to improve the discoverability of their content). Other common applications include using AI to identify market trends and support recommendation platforms.
- To date, **two-thirds of these publishers have realised benefits from their AI investments** and within two years, all publishers expect to have realised benefits. Current examples of realised benefits include improved IP protection, risk management and competitive advantage and receiving additional strategic insight.

## 3.1 Investment levels

Although it is still early days for the majority of publishers' investment in AI, our evidence suggests that both publisher interest in AI and year-on-year investment rates are increasing significantly.

Unsurprisingly, given the sums of investment required, large publishers (those employing at least 250 FTE staff globally) are leading investment activity. This is taking different forms. Some publishers have invested in dedicated small internal AI research teams, others have integrated external AI research and investment into their existing data and analytics teams, and some have acquired AI tech start-ups to quickly 'step up' the learning curve.

### 3.1.1 When did investment activity begin?

Our evidence suggests that while a few publishers were investing in AI before 2015, the majority of investment activity has occurred since 2017 when the number of investing publishers doubled. Some publishers are planning to invest in the next two years, whilst a significant proportion (these are all micro-sized publishers) have no plans to invest (Figure 10).

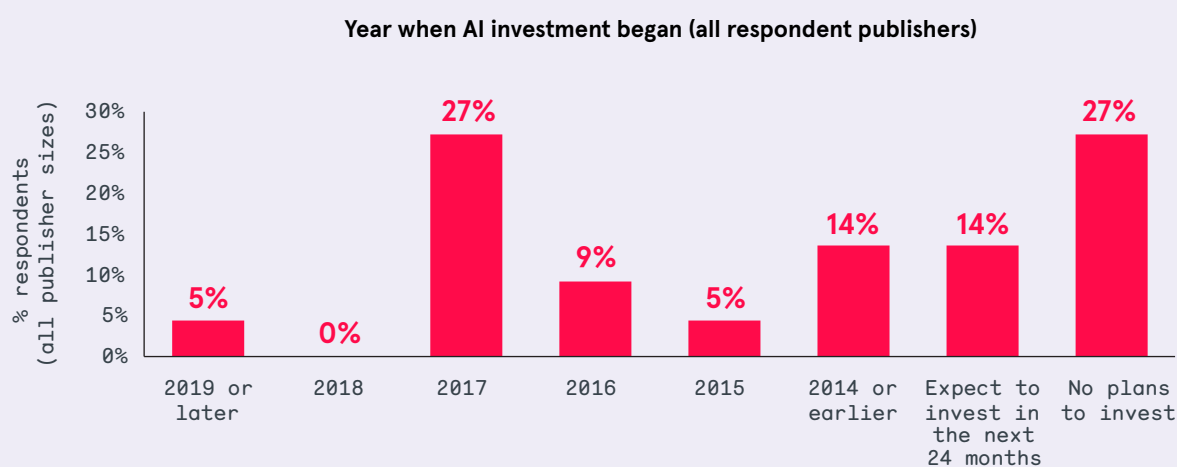


Figure 10. Source: Frontier Economics analysis  
Note: Base = 22 publishers (all sizes)

# 2017

The year when the majority of publishers started investing in AI

### 3.1.2 How significant is investment activity?

The publishers we interviewed indicated that a relatively small number of people within their organisation were focused on AI development or implementation. This is borne out by the industry survey responses. The majority of respondent publishers (all sizes) stated that they had five or less FTE staff researching or implementing AI technology within their global organisation.

Large AI-active publishers (see definition below in section 3.2) have on average slightly larger teams: around half of these respondents have five or fewer FTE staff researching or implementing AI technologies, and the majority of remaining respondents have teams of up to 20 FTE staff focussed on AI research.

Our survey also found that just under half of the staff that publishers reported as being in AI-specific roles were based in the UK.

We asked publishers to estimate how much they had invested in AI technologies over the last two years. Feedback suggested that many found it hard to distinguish the value of AI investments from wider investments – for example, the ‘AI component’ of an investment is often part of a wider technology research project, or closely linked with other technologies.

A total of 12 publishers (of all sizes) were able to estimate this for our survey. The majority of these respondents had invested up to £500,000 over the last two years. The remaining respondents indicated they had invested up to £5 million over the last two years.

The infographic features the text '1 to 5' in a large, bold, sans-serif font. The numbers are filled with a vibrant, abstract pattern of pink, purple, and blue. The letter 'o' in 'to' is replaced by a white silhouette of a person's head in profile, facing left. The background is dark blue with a network of glowing green and yellow lines and dots, suggesting a digital or AI theme.

The number of FTE staff carrying out AI-related tasks in the majority of respondent publishers' organisations

### 3.2 What is the nature of AI investment?

The survey evidence shows that the large publishers in our survey sample are responsible for the majority of AI investment activity to date. This is supported by evidence from the stakeholder interviews (discussed in Section 3.1) that suggests large publishers overall are leading investment activity.

As discussed in Section 1.4.3 our sample of large publishers represents a large proportion of both PA's large members and all large UK publishers. However, as we only received responses from those already engaged in AI investment we can only confidently interpret our results as being representative of large publishers that are already engaged in AI research or investment activities.

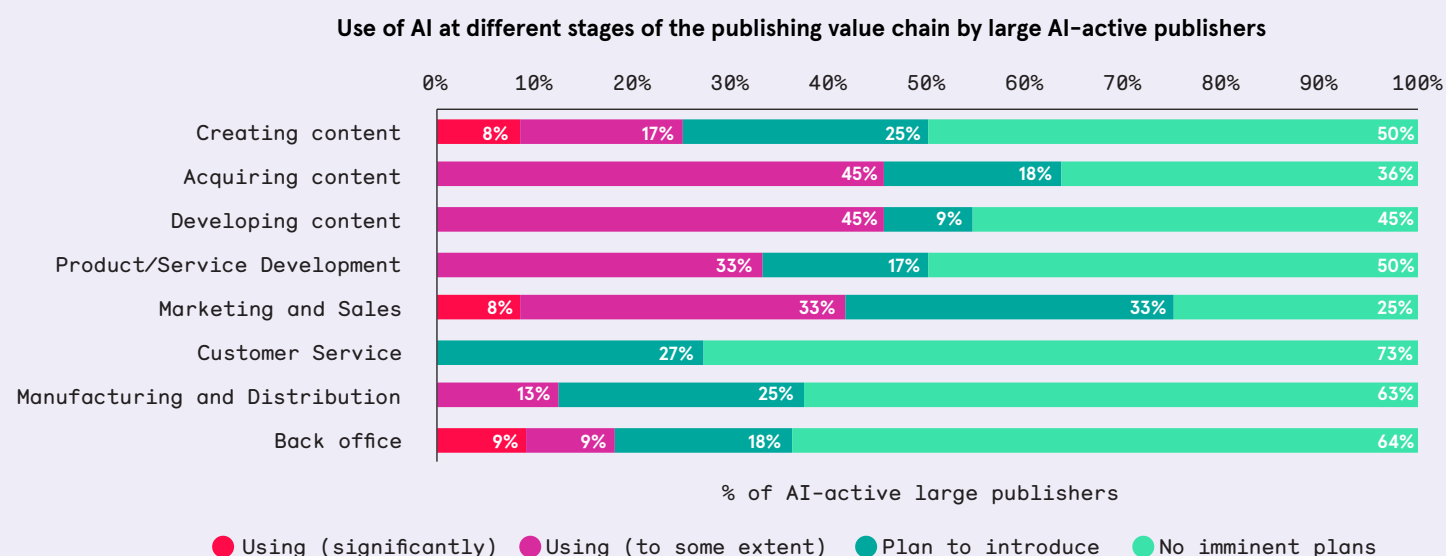
We refer to these large publishers that are investing in AI as ‘large AI-active publishers’. The survey evidence presented in this sub-section is based only on these 12 publishers’ responses.

## LARGE AI-ACTIVE PUBLISHERS

Publishers employing 250 or more staff globally that are using or researching how to use AI within their organisation.

### 3.2.1 Evidence on how large AI-active publishers are using AI

Our survey asked publishers to state where AI was already being used (or planned to be used within the next 2 years) in their value chain. Figure 11 presents the responses focusing only on the responses by large AI active publishers. Their responses are categorised in line with the taxonomy value chain stages (Section 2.1.1).



**Figure 11.** Source: Frontier Economics analysis  
**Note:** The base for this chart varies as publishers were allowed to leave a blank response or state not relevant for individual parts of the value chain. The maximum number of respondents was 12, and the minimum number of respondents was 8.

The results show that AI is currently being applied to all parts of the publishing value chain, except for customer service. Content acquisition and development are the most common parts of the value chain seeing current AI applications, though fewer than half of large AI-active publishers report current application of AI in these areas. The results are consistent with AI still being at a relative infancy in publishing.

There is little difference in the pattern of responses received from academic, education and consumer publishers. This was supported by our stakeholder interviews which captured AI use-case examples throughout the value chain for all sub-sectors

(although in some cases the precise application of AI differs reflecting differences in, for example, editorial processes and customer requirements – these differences are explained through examples in Sections 2.2.1 and 2.2.2).

### Where will AI be applied in the near future?

Looking forward to the next two years, there is evidence that AI will be used by a significantly greater proportion of publishers across different parts of the publishing value chain. For example, around a quarter of large AI-active publishers plan to introduce applications of AI in content acquisition, customer service and manufacturing and distribution in the next two years.

Overall, the survey results shown in Figure 11 and the literature and stakeholder evidence suggest that by around 2022, the most prevalent uses of AI may be with regard to:

- **marketing and sales activities:** this may include applications such as using AI to make customer/researcher content recommendations and to forecast demand.
- **publisher content acquisition:** this may include using AI to assist with market research, commissioning activities and to repurpose existing content.
- **content development:** this may include applications such as using AI to assist with proof reading, peer reviewing and editorial tasks.

We explore the potential for AI to generate new published content through our first case study (see box below) which considers one of the first examples of published AI-generated content, Springer Nature's (an academic publisher) Lithium-Ion batteries research publication.



# CASE STUDY 1

## AI and academic content generation (Springer Nature)

Springer Nature published ‘**Lithium-Ion Batteries: A machine-generated summary of current research**’ in 2019. The book, which took 1.5 years to develop and produce, represents one of the publishers’ most significant AI- projects to date. It is an example of publishers collaborating with AI and data experts to bring in the required AI expertise: the project brought together the expertise of Springer Nature’s Data Development team and their global area experts (chemistry and materials science), with the University of Frankfurt, and Digital Science (a science service partner).

The project is also an example of AI investment being **benefit driven**. The key motivation for producing the book was based on Springer Nature’s own research which showed that ‘information overload’ is a significant problem for researchers: for example there is no available subject overview that a person can run through in a couple of days to understand a problem. Another insight that motivated the project was their understanding that modern problems – such as meeting the UN’s Sustainable Development Goals require **inter-disciplinary working**, rather than siloed working practices. AI has the ability to cut through disciplines.

As well as these ‘user benefits’, the project intends to generate **industry spill-over benefits**. All of the code used to create the book has been placed in open-source enabling researchers to continue to develop the technology. Springer Nature hopes that this will lead to a greater level of information sharing to innovate more efficiently and to create a standard that will ultimately be accepted by various communities and markets.

The book itself was generated by applying AI technologies (a combination of machine learning and NLP technologies) to an initial long list of 1,086 publications identified through keyword searches. A simplified overview of the book generation process and key technologies employed is as follows:

- **Document clustering and ordering:** First, clustering technology was used to identify the specific contribution and scope of each input document, with a selection process used to prioritise and order documents by relevance.
- **Extractive summarisation:** Next, auto-summarisation technologies were employed which create excerpts from the documents and form the basis of the book’s subsections.
- **Paraphrasing and the generated extracts:** Finally, aggregation and paraphrasing techniques were used to improve the readability of the content.

The approach adopted by Springer Nature is considered conservative in order to preserve the audit trail. At any time, the reader is able to link through to the underlying source (something that was monitored by Springer Nature closely). ‘Black box’ AI solutions need to be considered and applied carefully by publishers, especially in the areas of peer review where audit trails are fundamental to judging quality. Similarly for scientific research, researchers require accurate reporting with a full audit trail. Springer Nature hopes that its first output in this area initiates a critical starting point for continuing research.

Source: Frontier stakeholder interviews and Beta Writer, 2019. Lithium-Ion Batteries: A machine-generated summary of current research.

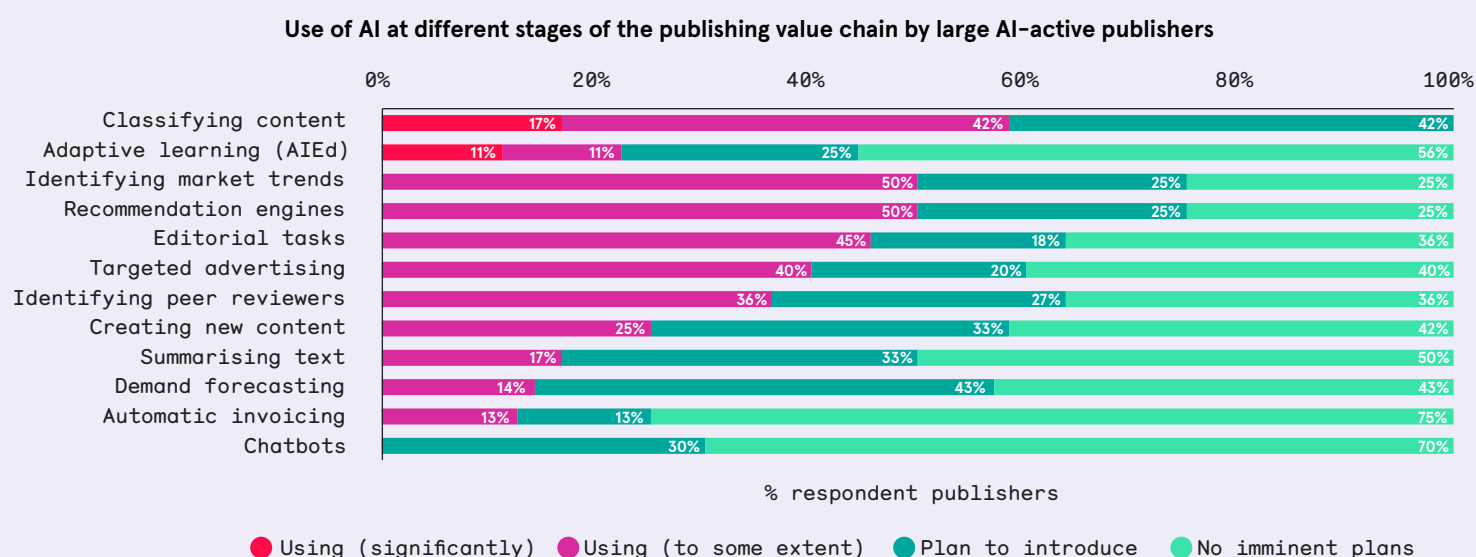
Our stakeholder evidence suggests that it may take longer to develop market-ready AI-generated content for consumer and education publishing. While there are already examples of AI-produced consumer novels, the consensus of those we engaged was that it will be a long time before AI is able to produce content that substitutes for human created content.

“1 The Road” was published in 2018 and marketed as the first published novel to be completely written by AI. It is described by its human author, Bruce Goodwin, as “very much an imperfect document, a rapid prototyping project” who also commented that it was far from being a human novel.<sup>31</sup>

In early 2020, a short-form novel “The day a computer writes a novel” written by a Japanese University research team, succeeded in passing the initial screening round of a literary prize.<sup>32</sup> The release in 2019 by OpenAI (a non-profit lab) of its Generative Pretrained Transformer 2 (GPT-2) technology is considered to be a step-forward for AI’s capability to generate creative content (including ‘deep fakes’).<sup>33</sup> Again, though, the time when GPT-2 or other AI technologies are able to ‘write a novel’ appears to be some way off<sup>34</sup>. This ties in with our evidence from consumer publishers and their plans to invest in AI-generated content.

## 3.2.2 Which tasks are being enhanced by AI?

We asked publishers about the specific tasks they are using (or plan to use) AI for. These results are shown in Figure 12.



**Figure 12.** Source: Frontier Economics analysis  
 Note: The base for this chart varies as publishers were allowed to leave a blank response or state not relevant for individual technologies. The maximum number of respondents was 12, and the minimum number of respondents was 7.

<sup>31</sup> As reported in <https://singularityhub.com/2018/10/25/ai-wrote-a-road-trip-novel-is-it-a-good-read/>

<sup>32</sup> [https://www.vice.com/en\\_us/article/wxnjn/a-japanese-ai-almost-won-a-literary-prize](https://www.vice.com/en_us/article/wxnjn/a-japanese-ai-almost-won-a-literary-prize)

<sup>33</sup> <https://openai.com/blog/gpt-2-1-5b-release/>

<sup>34</sup> <https://www.wired.com/story/nanogenmo-ai-novels-gpt2/>

AI is being used to some extent across all key task areas by large AI-active publishers. Based on the survey responses, we identify these four groups:

- **Common AI-assisted tasks** - currently carried out by at least half of the large AI-active publishers responding to our survey.
- **Moderately common AI-assisted tasks** – currently carried out by between a quarter and a half of large AI-active publishers responding to our survey.
- **Tasks that may become common in the future** – tasks not currently considered common or moderately common that at least half of AI-active large publishers expect to be using AI to carry out in the next two years.
- **Strand specific AI-assisted tasks** – currently carried out by at least half of the large AI-active publisher respondents in a particular sub-sector (academic, education or consumer).

### Common AI-assisted tasks

**Content classification** is the most common AI-assisted task amongst AI-active publishers. It is also one of the two tasks where a few publishers indicated they were using AI significantly. The academic stakeholders we interviewed told us that the most common reason for applying AI in this way was using meta data tagging to improve the discoverability of their content through key word searches, and to automate the classification of their content. There was no clear evidence from the survey that AI-assisted content classification is being used to a greater extent in some strands of publishing than others.

Other common AI-assisted tasks include **identifying market trends** (survey respondents provided examples of using AI to model time series trends, identify hot topics and conduct sentiment analysis) and **recommendation engines** (some stakeholders we interviewed were using AI to recommend content to academic researchers based on their research interests and download history). We provide an example of AI driven recommendation engines in our second case study below.

These findings are corroborated by our stakeholder conversations. Some large academic publishers told us that the focus of their initial AI investments is on areas that add value while not disrupting current organisational practices. They felt that tasks applied to finished outputs (content classification, recommendation engines) or to wider market analysis would sit within this group, whereas tasks specific to particular ‘editorial silos’ might be particularly disruptive to current practices.

# CASE STUDY 2

## AI and publisher content recommendation (Elsevier and Taylor & Francis)

Elsevier's Mendeley Suggest and Taylor & Francis's partnership with UNSILO illustrate the value of using AI to help academic researchers identify relevant research that will feed into future published research outputs.

The focus of Taylor & Francis's three-year partnership with UNSILO (a Denmark-based technology solutions company - specialising in providing publishing solutions - and part of Cactus Communications) is around generating meta-data for Taylor & Francis's content. This has enabled Taylor & Francis to create a recommendation engine using UNSILO's Recommend product. This **recommendation engine** is provided to academic library customers to help their users identify relevant Taylor & Francis content. It is also improving the quality of the meta-data it generates for content that Taylor & Francis supplies to indexing companies such as Google Scholar. In this case, AI is used to read academic content (using NLP) and subsequently creates a knowledge base to identify the most useful keywords that can be used as tags. There is a feedback loop enabling the algorithm to be fine-tuned.

Although Taylor & Francis does not formally track the direct benefits of its AI investment in these areas, it expects that this higher quality data is enabling researchers identify more relevant content in a shorter time period, and Taylor & Francis to increase sales volumes by improving discoverability.

Elsevier acquired Mendeley (a free reference manager and academic social network) in 2013. The **Mendeley Suggest** feature within the platform uses ML to generate recommendations for users on what to read and who to collaborate with. ML analyses data on the interests of similar users (explicitly stated or inferred from papers they are reading). A process of Collaborative Filtering then enables Mendeley to predict user preferences for a set of items based on past experience. The diagram below is a simplification of the most commonly used collaborative filtering algorithm, as an example of the process.

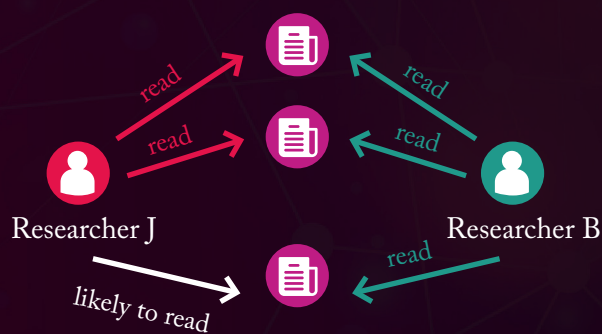
### Inputs

- There are N Researchers.  $R = 1...n$
- There are P Research Papers.  $P = 1...m$

How do we know if Researcher j ( $R_j$ ) will like paper c ( $P_c$ ) if they haven't read it?

Fine users similar to  $R_j$  and see if they liked the paper

### Collaborative filtering



Source: Frontier Economics analysis

Mendeley's own use case studies provide evidence of a range of user benefits, including helping researchers from ETH-Zurich and ETH-Bibliothek to collaborate across 'boundaries'<sup>35</sup>, helping the Institut Pasteur discover new research,<sup>36</sup> and helping the International Food Policy Research Institute to track and promote its authors' papers.<sup>37</sup>

<sup>35</sup> [https://www.elsevier.com/\\_data/assets/pdf\\_file/0019/107434/MendeleyCustomerSpotlight\\_ETHZurich\\_Final.pdf](https://www.elsevier.com/_data/assets/pdf_file/0019/107434/MendeleyCustomerSpotlight_ETHZurich_Final.pdf)

<sup>36</sup> Source: [https://www.elsevier.com/\\_data/assets/pdf\\_file/0020/107435/MendeleyCustomerSpotlight\\_InstitutPasteur\\_Final.pdf](https://www.elsevier.com/_data/assets/pdf_file/0020/107435/MendeleyCustomerSpotlight_InstitutPasteur_Final.pdf)

<sup>37</sup> Source: [https://www.elsevier.com/\\_data/assets/pdf\\_file/0004/69673/IFPRI\\_case\\_study\\_v5.pdf](https://www.elsevier.com/_data/assets/pdf_file/0004/69673/IFPRI_case_study_v5.pdf)



### Moderately common AI-assisted tasks

AI is being employed by publishers to a more moderate extent to help develop (edit) their output. Just under half of the large AI-active publisher respondents stated they were using AI in their **editorial tasks** (one survey respondent provided the example of a copyright infringement tool; another is using AI to conduct plagiarism checks). A smaller proportion of AI-active large publisher respondents stated they were using AI to help **identify peer reviewers**. This was highlighted by a couple of survey respondents as a key example of how AI is benefitting their business.

Interestingly, one in four AI-active large publishers reported that they are already using **AI to help create content**. Although these publishers did not provide specific examples, these are all active in academic publishing which fits with our understanding from the literature and stakeholder evidence that AI is being deployed most commonly in academic publishing to generate new content.

### Tasks that may become common in the future

The evidence on where AI will be deployed for the first time by publishers over the next two years suggests there will be a large increase in the number of publishers using AI to **classify content** (within two years all large AI-active publishers said that they expect to be using AI in this way), and to **forecast demand** (a significant proportion of AI-active large publishers stated they expected to start using AI in this way within two years). A few AI-active large publishers provided examples of using AI to forecast physical stock levels for their warehouse and to assist with their production processes as an example of applications in this area.



of large AI-active publisher respondents expected to be using AI to classify their content within 2 years

### Strand specific tasks

In the survey, six large publishers active in educational publishing told us about their current or planned use of AI in **adaptive learning**. Four of these indicated they were using or plan to use AI to provide adaptive learning solutions – this is the focus of our third case study, below. The case study suggests that adaptive learning has the potential to deliver significant user benefits. The case study also highlights the importance of collaboration with organisations outside of publishing to bring this technology quickly to market.

These findings are backed up by the survey evidence. Three of the seven responding AI-active large education publishers stated they were largely dependent on external organisations for acquiring AI technology and expertise, and only one stated they had no dependence on external organisations. As a result, publishers are facing increasing levels of competition in this area from non-publisher EdTech companies. Half of publishers active in education publishing in our survey expect to face additional competition from tech companies.

# CASE STUDY 3

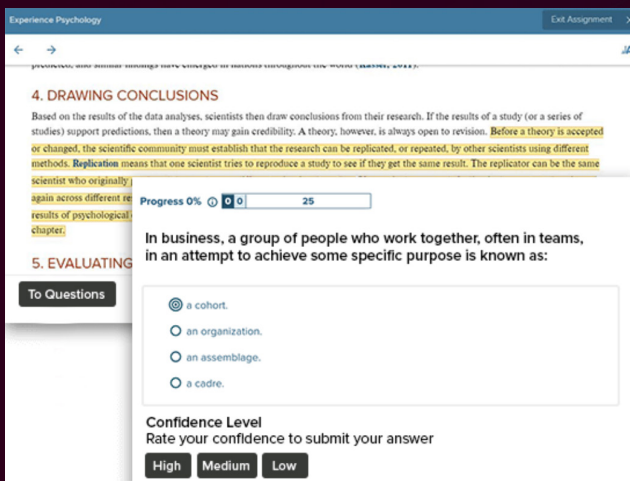
## AI and education publishing product development (McGraw Hill)

McGraw Hill, an education publisher, has invested significantly in AI technology. Its two AI-assisted platforms, ALEKS and SmartBook, exemplify how AI is driving significant wider benefits to educators and learners.

The ALEKS platform (‘Assessment and Learning Knowledge Spaces’) was developed in partnership by a team at the University of California and McGraw Hill, with the help of a grant from the US’ National Science Foundation. The partnership resulted in a platform that uses AI to deliver tailored content to maths, science and business school and Higher Education students. McGraw Hill acquired the ALEKS Corporation in 2013.

The mapping is carried out using Knowledge Space Theory. This draws on the idea that there are many possible states of knowledge of a human learner. AI is applied (using a series of 20 to 30 questions) to understand precisely which part of the knowledge space a student is in. This information determines which topic to teach the student next.

SmartBook, McGraw Hill’s other AI platform, was also developed externally by Danish EdTech company, ‘Area9’, and acquired by McGraw Hill in 2014. SmartBook is an adaptive reading tool. The latest iteration was launched in 2019. It augments textbook content by highlighting the most relevant text for a student and uses AI to assess students’ knowledge and guide them in reinforcing this.



### SmartBook

This is a snapshot view of SmartBook showing the text and supporting text question that a user would see.

- Yellow highlights are used to indicate key text passages
- The test is a multiple-choice question, and a user confidence rating

McGraw Hill cites student benefits from these platforms as including increased engagement, improved study time effectiveness and higher quality learning (learning gradually rather ‘cramming’) resulting in deeper understanding and improved recall rates. Student feedback to McGraw Hill is positive. Educators benefit from saving time and so can use their classroom time more effectively. They can also access improved student progress analytics, enabling them to quickly spot trends that may take longer to feed through in larger classes, or quickly identify struggling students. The platforms are being used as a recruitment tool – for example, Brunel University promotes the SmartBook platform at Open Days.

Source: Frontier Economics stakeholder research



### 3.2.3 What are the benefits of employing AI in publishing?

The final part of our taxonomy sets out the potential for AI to deliver benefits to publishers and the wider economy. Publishers told us that these benefits are a key driver in their decision to invest: understanding these benefits provides an indication of future investment intentions.

There are a number of ways to measure these benefits. Our survey analysis provides:

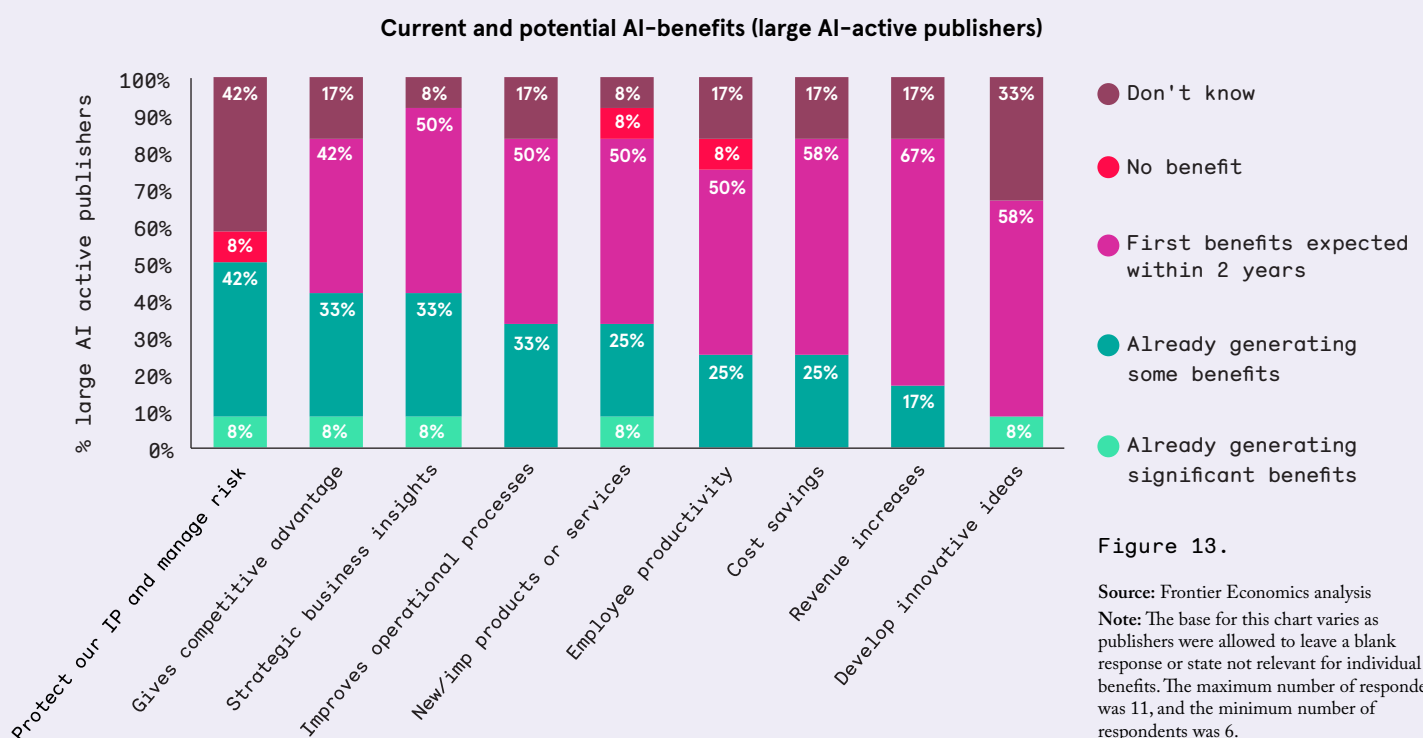
- Evidence on the extent to which publishers are realising (or expect to realise) benefits from their AI investments; and
- Evidence on the type of benefits publishers have realised (or anticipate).

#### The extent to which publishers are benefiting from AI

The survey responses suggest that the majority of large AI-active publishers have realised AI-related benefits. Two thirds of large AI-active publishers reported that their AI investments are already generating benefits.

Within two years, all large AI-active publisher respondents said that they expect to be realising some AI-related benefits.

Based on the taxonomy developed for this study, we asked publishers which types of benefits were realised or anticipated. 12 large AI-active publishers responded. Of these, one had already realised significant benefits and a further eight had already realised some benefits. The results are summarised in the chart below for the large AI-active publisher group.



Half of this group reported they were already benefitting in terms of being able to **protect IP and manage risk** (for example, by using AI to detect plagiarism). More than two in five also reported benefits in terms of **provided competitive advantage** and **providing strategic insight**. However, consistent with the relative maturity of AI investments in publishing so far, few publishers reported ‘significant’ benefits to date.

Looking forward, publishers are optimistic about the **benefits that AI will deliver within two years**. Figure 13 also shows that in the next two years the proportion of AI-active large publishers benefiting from AI will **significantly increase in all but one of the benefit areas** we asked them about (‘protect our IP and manage risk’ – this is also the main benefit area where publishers were unable to express a view).



# 4



## The barriers to AI investment

The stakeholder interview evidence identified a range of potential AI investment barriers. We explored these further with all publishers through the survey. This included the 12 large AI-active publishers as well as the micro, small and medium publishers, the majority of whom are not yet investing in AI.

In the survey, publishers were asked to consider a long-list of barriers that could inhibit their ability to invest in AI and indicate which of these they were either currently facing, or expected to face in the next two years. Publishers could indicate whether the barrier was expected to be significant, slight, potential or unlikely. Next, publishers were presented with a short-list of barriers which had been drawn up in consultation with publisher stakeholders. In this case, publishers were asked to identify their most significant barrier.

Long list (survey wording)		Short list (survey wording)
Ability to acquire skills	Awareness of potential benefits (of AI)	Data availability
AI skills within the business	Lack of leadership support for AI	Current IT infrastructure
Current IT infrastructure	Current organisational structure	AI technology availability
Data availability	Resistance to changing working practices	Organisational issues
AI technology availability	Current government policy/regulation	AI skills availability
AI research/acquisition/implementation costs		Policy and regulation

Figure 14. Source: Frontier Economics

Our analysis of the survey results revealed significant differences between the barriers that large AI-active and the SME publishers in our sample are facing. We present our findings for these two groups separately and then highlight the key differences.

KEY FINDINGS

Our survey evidence suggests that a lack of AI-skills and difficulties applying AI solutions to existing IT infrastructures are the most common AI investment barriers faced by large AI-active publishers. To overcome the skills issue our survey and case study evidence shows that many publishers are collaborating with external research organisations (half of the large publishers we surveyed on this issue consider themselves largely dependent on external expertise for acquiring AI skills and technology).

Lack of awareness of the potential benefits of AI also appears to be a significant barrier for large AI-active publishers. We understand this may be linked to a lack of scientific knowledge by investment decision-makers and may also be attributable to the fast rate of technological change currently seen for AI.

Organisational barriers were mentioned by several stakeholders. The survey evidence suggest this is a slight barrier for the majority of large AI-active publishers. The stakeholder interviews and case study evidence suggests that reducing other barriers on awareness of potential benefits and internal AI-skills (particularly decision-makers’ understanding of where AI can add benefits) could be an effective way of addressing this barrier.

While we have limited evidence on SME publisher barriers, it appears that the large upfront costs associated with researching and implementing AI solutions may be prohibitive for smaller publishers.

4.1 Large AI-active publisher barriers

Figure 15 summarises the survey responses from large AI-active publishers.

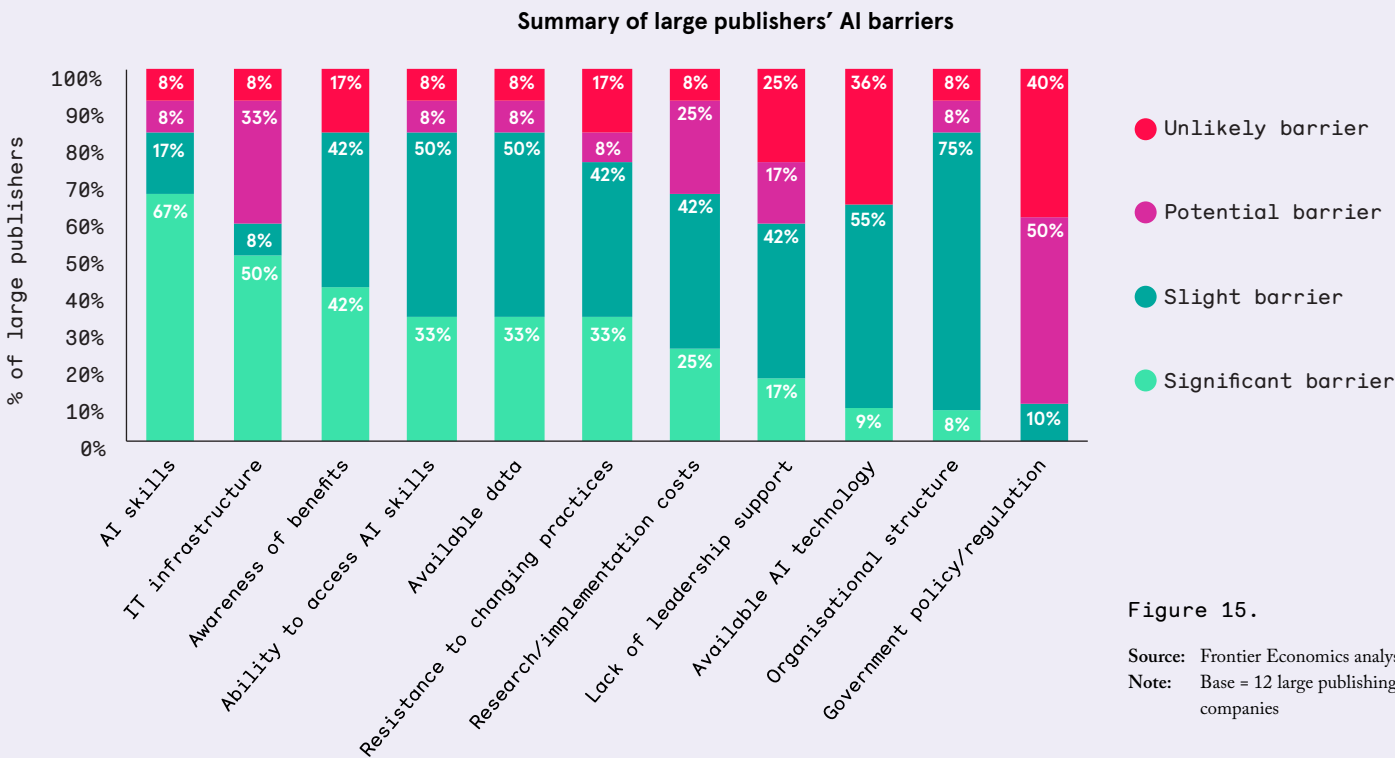


Figure 15.

The chart identifies where publishers have indicated that a particular barrier is current (significant or slight); a potential barrier; or unlikely to be a barrier.

We discuss our findings for current and future (the sum of current and potential) barriers and then bring in additional evidence to discuss publishers' most significant barrier.

### 4.1.1 Current barriers

**AI skills within the business** are currently the most significant AI investment barrier for our sample of large AI-active publishers. Two thirds of respondents cited this as a current, significant barrier. One respondent told us that their organisation needed to acquire more scientific and technical skills and most publishers are collaborating with AI-tech firms to gain these skills (see box below).

The next most significant barrier appears to be **current IT infrastructure**. Half the respondents in this group said this was a current, significant barrier. One stakeholder described the problems currently faced moving ideas out their AI-research lab an applying these to the wider IT infrastructure.

**Awareness of potential benefits** is the third most significant barrier for our sample of large AI-active publishers. Many stakeholders told us that it was challenging to get buy-in across the organisation because many decision-makers had a low awareness of the benefits. To address this some publishers told us they were investing in AI in the least disruptive areas of the organisation in order to provide evidence. Other stakeholders linked this barrier back to the AI skills barrier, stating that investment decision makers often lack the scientific skills to understand the potential benefits of AI, and that there was a 'misunderstanding of where machine learning could be used and the areas where the most benefit could be gained'. One survey respondent told us simply that given the current rate of technological change it was 'hard to predict where to place the bets'.

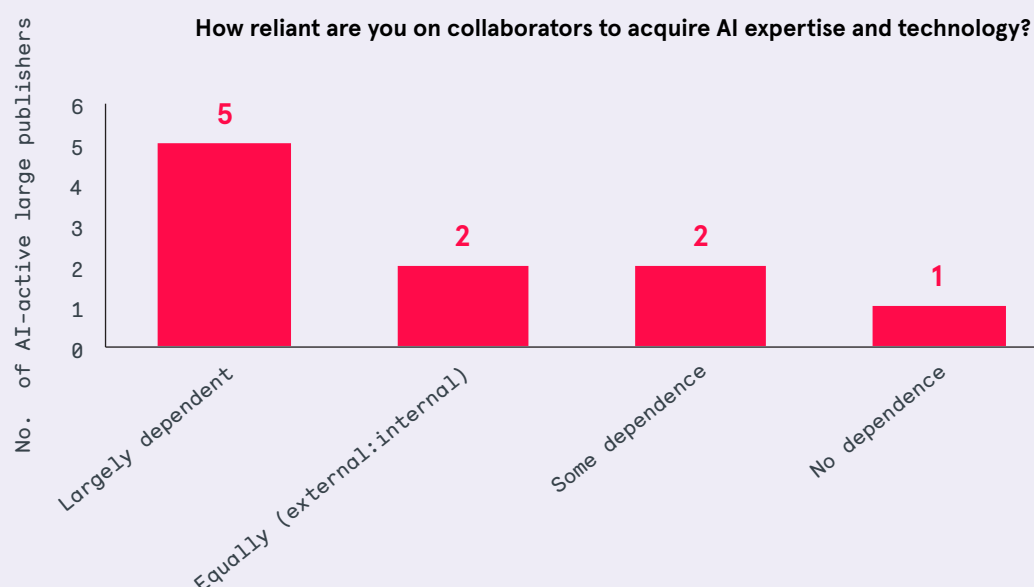
## OBTAINING AI SKILLS THROUGH COLLABORATION

10 of the 12 large AI-active respondents had engaged with external organisations to acquire AI expertise and technology within the last two years (another publisher stated they are investigating a future partnership). The chart below shows/ demonstrates that AI tech companies are the most popular type of collaborator, but large publishers had also engaged with consulting firms, universities, EdTech firms and incubators or accelerators (publishers were able to select more than one type of collaborator).

External organisations are an important source of knowledge for publishers. Five of the ten publishers who had collaborated in this period consider themselves as 'largely dependent'. Our stakeholder interview evidence suggests this dependency is one way that publishers are handling the considerable investment risks associated with acquiring specialised technical knowledge in a period of significant change.

Our three case studies all provide examples of large AI-active publishers collaborating with non-publishing organisations to acquire AI expertise and develop innovative publishing outputs. In each case, the publishers acknowledge that their project's success can be attributed to being able to combine the strengths of their organisation with those of others outside of the industry.

- Case Study 1 describes **Springer Nature’s collaboration with the University of Frankfurt and Digital Tech** (a UK AI-tech company). The publisher particularly valued the ‘start-up’ culture brought by the tech company and the cutting edge developments from the University, and commented that the overall combination of different areas of expertise and research styles added strength to the project.
- Case Study 2 highlight’s **Taylor & Francis’s 3-year partnership with UNSILO** (a Danish AI-tech organisation specialised in publishing solutions, recently acquired by Cactus Communications). We understand UNSILO has also partnered with Cambridge University Press and Wiley.
- Case Study 3 features the McGraw-Hill ALEKS adaptive learning platform, developed through **a partnership between the University of California and McGraw Hill**. Here McGraw Hill attributes the success of this project to partnership working which combined scientific cutting-edge scientific expertise with the publishers’ expertise in developing content to meet customer needs. An AIED expert stakeholder we interviewed for this case study stated that in Education Publishing, many publishers are having to become tech companies and to do so are having to bring in significant external knowledge from tech companies. As a result there are an increasing number of Ed-Tech start-ups, which in time may disrupt the education publishing sub-sector.





## 4.1.2 Future barriers (current and potential)

Figure 15 shows that at least 90% of large AI-active respondents considered the following to be either current or potential future barriers:

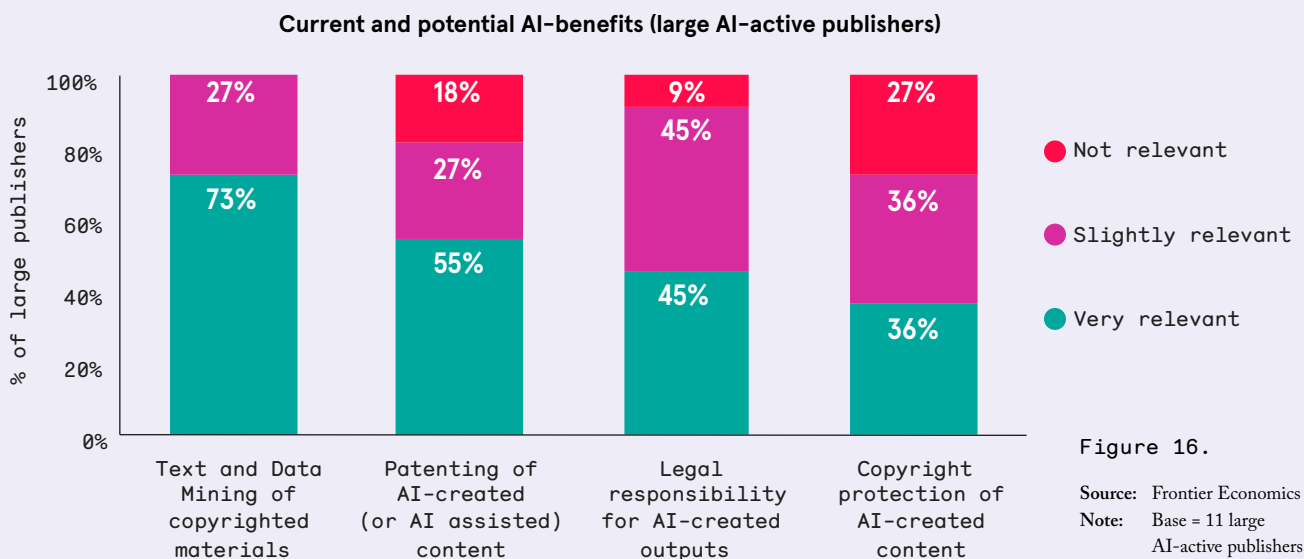
- AI skills within the business
- Current IT infrastructure
- Ability to acquire AI skills
- Available data
- AI costs
- Current organisational structure

Looking specifically at those who pointed to potential future barriers (but not necessarily current barriers), **government policy and regulation** and **current IT infrastructure** were most commonly-cited. Half of respondents said that policy and regulation were potential future barriers. A third of respondents that IT infrastructure was a potential future barrier – on top of the 58% who cited it as a current barrier. We explore publishers' concerns over future government policy and regulation below.

### Potential future policy and regulation barriers

Policy and regulation is not currently considered an investment barrier by large AI-active respondents. In Figure 15, only one in ten respondents indicated this to be slight barrier. In the future, large AI-active publishers appear to be more concerned about policy and regulation as an investment barrier with an additional five in ten publishers considering policy and regulation a potential barrier.

Our literature review and case study interviews identified some concerns about Intellectual Property (IP) policy. Our survey explored which aspects of Intellectual Property policy and regulation are most relevant to publishers' AI investment, using a short-list created from published material.<sup>38</sup>



We asked large AI-active publishers their opinion on how relevant four specific policy areas were to their business. These were: Text and Data Mining (TDM); legal responsibility for AI created outputs; copyright protection for AI created content; and patenting of AI created content.

The responses are shown in Figure 16. In summary, for those expressing a view:

- **TDM of copyrighted materials** (see box, below) is considered as ‘very relevant’ by nearly three-quarters of large AI-active publishers, and all indicated this was at least slightly relevant.
- **Patenting of AI created or assisted content** was considered very relevant by more than half of large AI-active publishers.
- The question over who assumes **legal responsibility for AI created outputs** was considered as very relevant for their organisation by just under half of large AI-active respondents.
- **Copyright protection of AI content** was indicated as very relevant by just over one in three publishers.

## THE KEY ISSUES CONCERNING TDM AND COPYRIGHT LAW

Creating a robust AI algorithm requires considerable volumes of data. For publishers this data can be sourced either from their own content, or from content where they do not own the rights.

TDM has the potential to create significant value as the ‘fuel’ for AI. The value that is being extracted by TDM activities comes from the patterns and the discovery of new relationships between the data, rather than the data or text itself. In June 2019, the EU adopted the **Directive on Copyright in the Digital Single Market**. This contains two mandatory TDM exceptions which allow:

- all research institutions (e.g., universities, libraries, hospitals, research institutions) and cultural heritage organisations to use copyrighted material to train their AI; and
- anyone else to use copyrighted material to train their AI, provided that rights holders have not opted out.

Member States have until 2021 to transpose this into national law although following Brexit, the UK has indicated that it does not currently intend to do so.

The UK already has its own TDM exception (since 2014), which allows researchers to make copies of copyright material for the purpose of computational analysis if they already have lawful access to the work, and provided the making of copies is for research of a non-commercial purpose.

---

38 For example, material published by IPKat: <http://ipkitten.blogspot.com/2019/06/ai-decoding-ip-conference-day-1.html>

## 4.1.3 Most significant barrier

11 large AI-active publishers selected their most significant barrier (taking into account current barriers and those anticipated in the next two years) from the short list in Figure 14. **Data availability** and **organisational issues** were most significant for large AI-active publishers (Figure 17).

Barrier	No. large AI-active publishers indicating this as most significant barrier
Data availability	4
Organisational issues	4
AI skills availability	1
IT infrastructure	1
No barriers	1
£20m+	0

Figure 17. Source: Frontier Economics analysis

### Data availability barriers

AI technology requires increasingly large volumes of data to train algorithms. Several stakeholders we interviewed viewed this as a constraint, suggesting that publishers are having to ‘catch up’ with producing data that satisfies the needs of emerging technologies. Large publishers are investing considerable resources currently to add metadata to their back catalogue, evidenced by our earlier finding that all large AI-active publishers expect to be employing AI to classify their content within the next two years (see Section 3.2.2).

### Organisational barriers

Stakeholders told us that implementing an AI solution represents a considerable undertaking for their organisation. As noted in Sections 4.1.1 and 4.1.2, training an AI algorithm requires significant investment in skills and IT infrastructure as well as the acquisition and preparation of the large volumes of data. We understand that many large publisher organisations have not undertaken IT infrastructure investments of this scale before, and that some are finding that the way their businesses are organised is creating a significant barrier to investment.

A number of academic publisher stakeholders explained that their organisational barriers derive from the way that the content acquisition and development parts of their value chain are organised in silos. Each silo is controlled by an editor, and operational practices can be quite different from one silo to the next. The cost of AI solutions means they have to be applied across a publisher’s organisation. A number of stakeholders we spoke to who were trying to develop solutions that cut across these silos said that trying to implement a common solution was currently being met with resistance. We noticed that in these cases publishers were tending to focus their investment activities on the non-siloed areas of their business. One stakeholder we interviewed stated that by demonstrating the efficacy of AI investment elsewhere in the business (such as developing recommendation engines – see Case Study 2), where

it is easier to implement cross-cutting solutions, it was hoping to demonstrate the benefits of AI to decision-makers within the siloed areas.

## 4.2 SME publisher AI barriers

Figure 18 summarises the survey responses we received from SME publishing organisations on the AI barriers they are facing, or expect to face in the next two years.

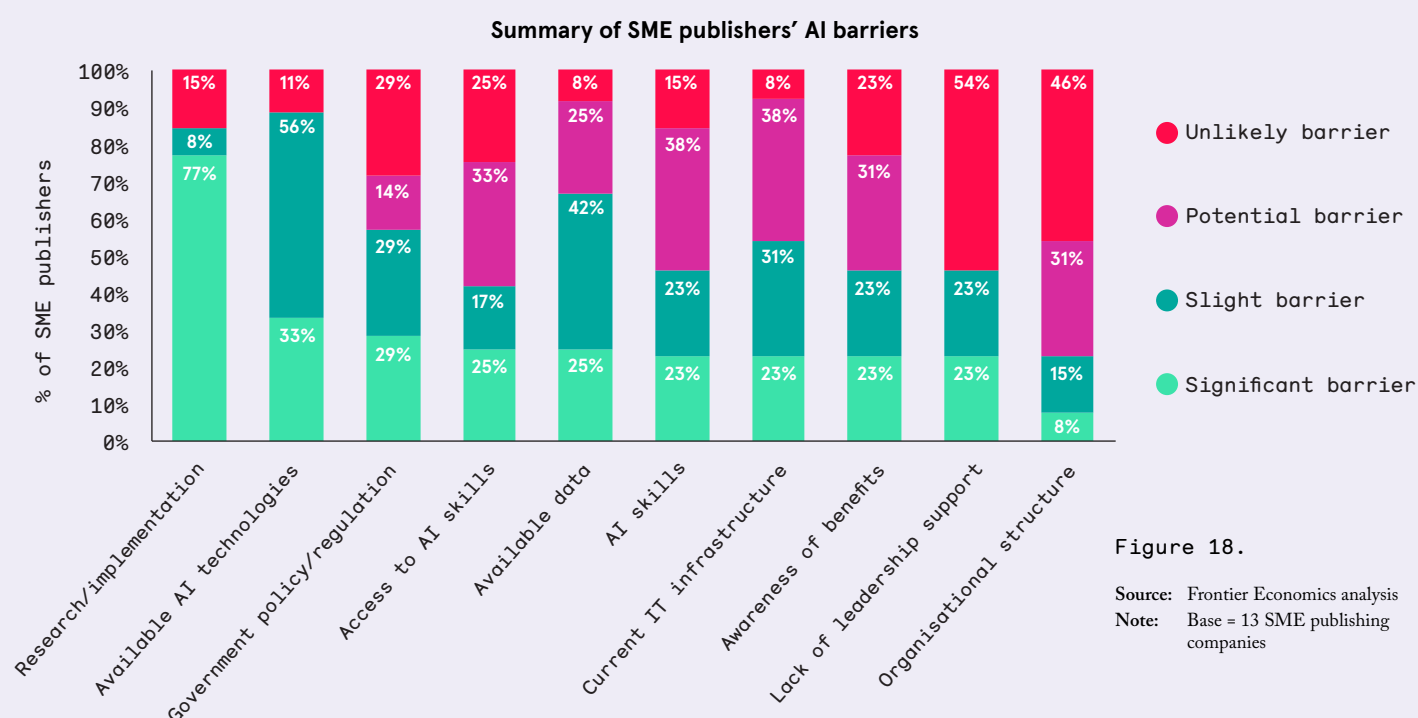


Figure 18 shows that **AI costs (researching/acquiring and implementing)** appear to be a significant barrier for most (77%) SME publishers. When we consider the wider characteristics of this SME publisher group this is not surprising:

- 13 of the 14 SME respondents stated they were not using or researching any form of AI, and 11 of the 14 SME respondents indicated that AI is not a priority investment for their business.
- The majority of respondents in this group employ fewer than 10 employees and have annual turnover below £500,000, suggesting limited internal resources for current AI investment.

Taking into account both slight and significant barriers, Figure 18 shows that **availability of AI technologies** is considered a barrier by almost 9 in 10 respondents. 11 of the SME publishers identified their most significant barrier from the shortlist of potential barriers. The most common choice was **AI skills availability** (4 publishers).

## 4.3 Comparing AI investment barriers for large AI-active and SME publishers

Figure 19 compares the responses we received regarding SME and large AI-active publishers' significant barriers.

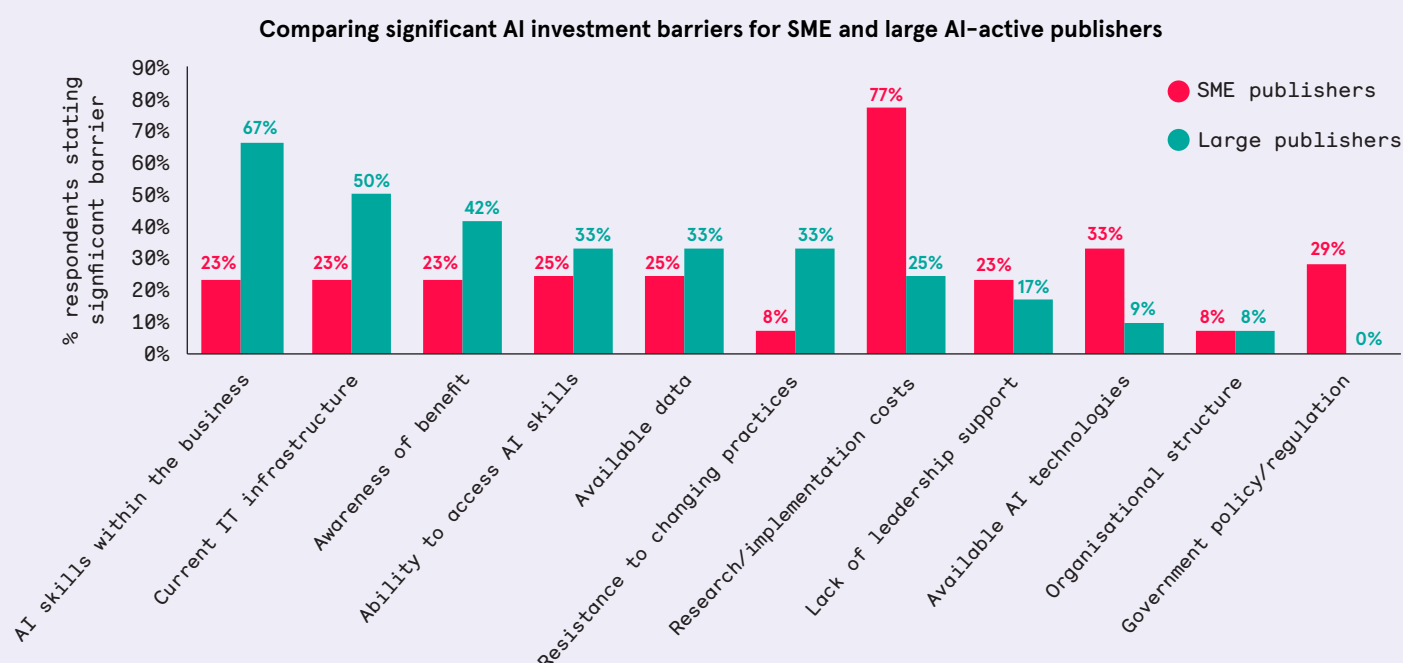


Figure 19.

Source: Frontier Economics analysis

Note: The base for each category excludes publishers who did not provide a response or who stated 'don't know'. As a result the base for each category is different. In this case the majority of categories are based on 12 large publisher and 13 SME publisher responses. The minimum number of responses is 7 for SMEs and 10 for large publishers.

Currently, the barriers faced by these two groups are quite different. Over three-quarters of **SME publishers** consider the **costs of researching and implementing AI solutions** as a significant barrier. Although this group identified other significant barriers, in each case a much smaller proportion indicated these (no more than one-third of SME respondents). This is in contrast with large AI-active publishers, the majority of whom do not consider costs to be a significant investment barrier (three-quarters did not indicate cost as a significant barrier).

**Large publisher** respondents considered **internal AI skills** (selected by two-thirds of publishers) and their **current IT infrastructure** (selected by half of large publishers) to be a significant barrier. The majority of SMEs do not consider these as significant current barrier (around one-quarter of SMEs identified these as significant barriers). It is possible that at present, cost barriers are so significant for SMEs, that other barriers, currently appear less significant in relation. SME publishers' views on potential barriers (Figure 18) suggest that in the future the barriers faced by SME and large publishers may be more similar (for example, an additional one-third of SMEs consider internal AI skills and current IT infrastructure barriers as potential barriers). As our SME survey statistics are based on a small sample we recommend further research in this area to better understand the differences in investment barriers faced by SME and large publishers.

# 5



## The importance of AI for publishing in the future and recommendations

Given the rapid rate of technological change for AI in publishing it is important that this report is forward looking. The stakeholders we interviewed and the wider literature evidence on AI in publishing<sup>39</sup> suggests that, in future, AI will play an important role in the publishing industry, both in terms of the extent to which it transforms publishers' organisations and also the extent to which it transforms the competitive landscape. We explore these issues below.

### KEY FINDINGS

The publishers we surveyed believe that **AI will be important for the industry over the next five years**. On a scale of 0-100, where 100 indicates high importance, publishers scored the importance of AI to their industry over the next five years as 69/100.

AI has the potential to significantly transform publishers' organisations. In our survey, one in six publishers expect to experience significant transformation and all large publishers expect AI to have at least a small impact.

It may take some time for this transformation to begin. In our survey, around half of large publishers expect transformation to begin within three years and the remainder within ten years. Transformation may take longer for SME publishers, one-third do not expect AI to transform their organisation and the majority expect this will occur in more than five years.

AI may have a significant impact on publishers' competitive environment. Our survey evidence suggests that almost two-thirds of all publishers (and eight in ten large AI-active publishers) expect to be competing with a different range of organisation types in the future. Our stakeholder interview and case study evidence identified AI-tech start-ups as potential future competitors. The survey evidence supports this, finding that 10 out of 12 publishers expecting to face new competition, expect this to come from SME tech start-ups and large established tech companies.

### 5.1 The future importance of AI for publishing

Publishers believe that AI will be of great importance to their industry in the next five years.

<sup>39</sup> For example, Gould Finch. 2019. The future impact of Artificial Intelligence in the publishing sector. Summary available at: [http://ifro.org/sites/default/files/colin\\_1911104\\_edinburgh.pdf](http://ifro.org/sites/default/files/colin_1911104_edinburgh.pdf)



In our survey, we asked respondents to rate, on a scale of 0-100, how important AI was expected to be for publishing in the next five years. The median response was 69.

There were some different perceptions between the large, AI-active publishers in our sample and other publishers. The median response for large, AI-active publishers was 72 and that of other publishers was 58. Large publishers rated the importance of AI in the next five years more highly (72/100) than SME publishers (58/100). There was also more variation in the responses for SME publishers (see Figure 20). However even among this group there were a large number of respondents who felt AI would be very important.

# 69/100

Publishers' average score for the importance of AI in publishing over the next five years

The importance of AI to the publishing industry in the next five years (0=unimportant, 100=very important)

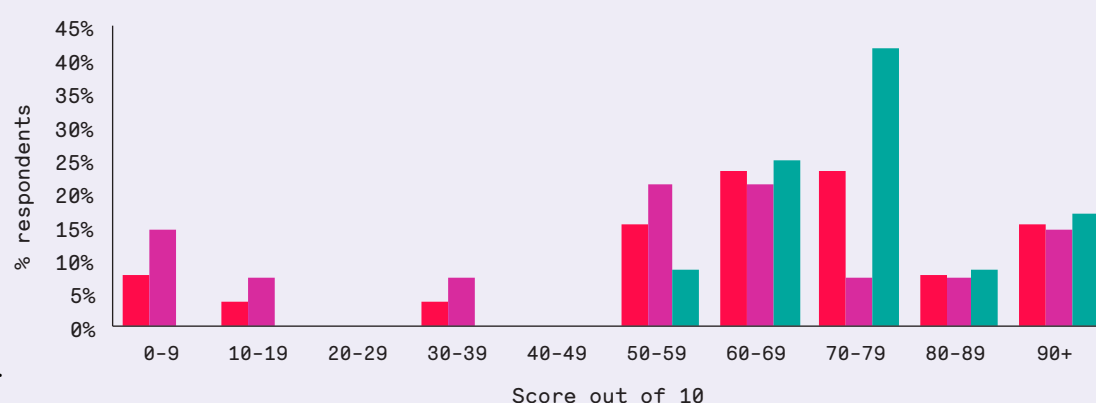


Figure 20.

Source: Frontier Economics analysis

Note: Base = 26 publishers  
(14 SME, 12 large)

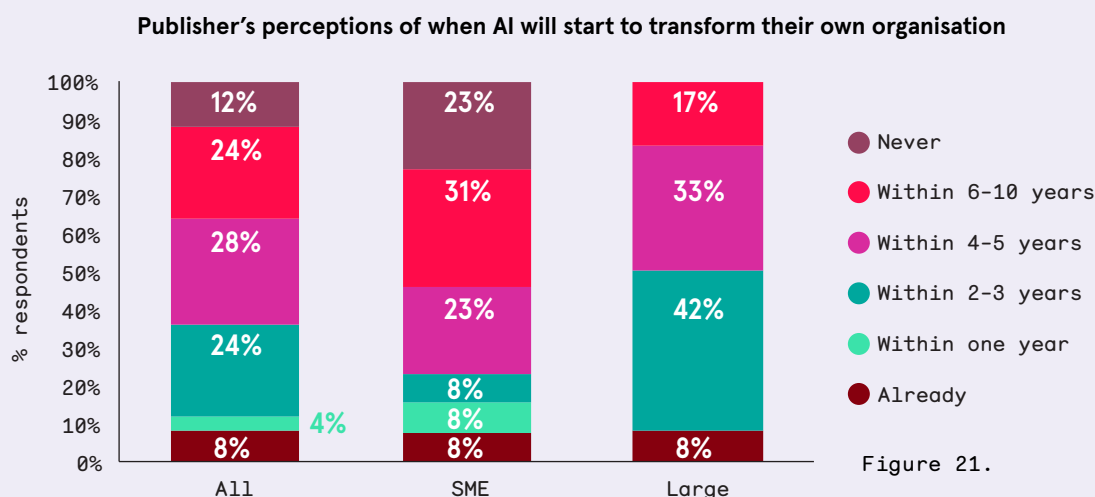
● All publishers ● SME publishers ● Large publishers

## 5.2 The potential for AI to transform publishers' organisations

Our survey asked publishers how quickly they expect AI to start transforming their organisation and the extent to which they expect their organisations to be transformed.

### 5.2.1 Timescales for AI transformation

Over 80% of respondents expected that AI will start to transform their organisation within the next ten years. Figure 21 shows responses and compares large AI-active publishers to SMEs.

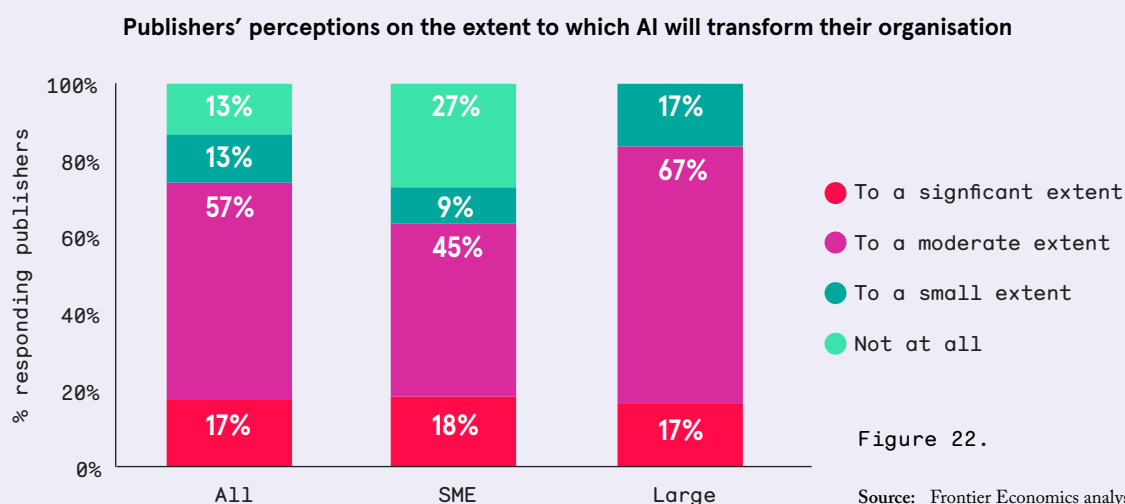


Whereas **all large (AI-active) respondents expect transformation to start within the next ten years**, around a quarter of SME publishers said that their own organisation would not be transformed.

Around half of large AI-active publishers expect their organisation to be transformed within the next three years.

## 5.2.2 Potential extent of AI transformation

Figure 22 shows that around three-quarters of respondents expect that AI will transform their organisation at least moderately, and one in six expect that this transformation will be significant. The evidence also suggests that large publishers expect to see a more significant transformation than SME publishers. Whereas just over one-quarter of SME publisher respondents stated they did not expect AI to have any impact on their organisation, all large respondents expected AI to have at least a small impact.



## 5.3 The potential for AI to transform publishers' competitive landscape

Many of the publishers we interviewed spoke of the potential for AI to transform their competitive landscape. This effect stems from their need to rapidly acquire AI skills and the role of non-publishing organisations (such as AI-tech companies) in meeting this need.

In our survey, nearly two-thirds of all respondents and over eight in ten large, AI-active respondents indicated that they expect the type of organisations they compete with to change as a result of AI investment in the publishing industry. Large publishers expect AI have to a greater impact on the nature of their competitors than SME publishers: around eight in ten large publishers expect to see a change, compared with less than four in ten SME publishers.

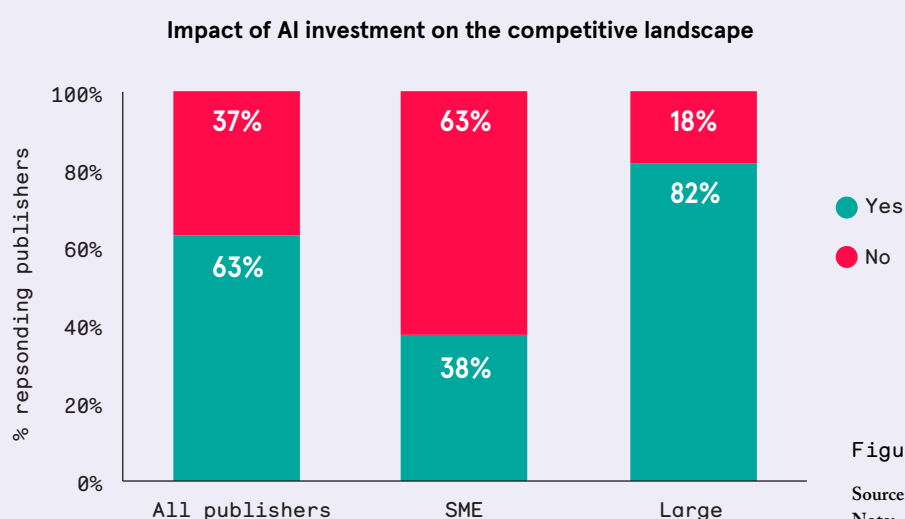


Figure 23.

Source: Frontier Economics analysis

Note: Base = 19 respondents (8 remaining respondents stated 'don't know')

Our stakeholder and case study evidence suggests that large AI-active education publishers are already facing new forms of competition. An AIED expert we interviewed as part of the AIED case study (Case Study 3) told us there were an increasing number of EdTech companies starting up. Currently the majority of these are acquired by big tech companies or large publishers, but there is potential for these to start to grow in their own right and disrupt education publishing.

Our survey evidence supports these findings. Of the 12 publishers<sup>40</sup> who stated they expected that AI investment by other companies would lead to their organisation facing new competitors, 10 expect to face new competition from SME tech start-up companies and established large tech companies. 5 publishers expect to face new competition from established SME tech companies.

<sup>40</sup> 9 of these are large publishers, 1 is small and 2 are micro.

# 6

## Recommendations for supporting future AI investment by publishers

This report presents evidence on the potential for AI investment to generate significant benefits for the UK publishing industry, the wider supply chain, and wider society.

A key insight is that in publishing, the benefits of AI for the industry appear to be about reducing time spent on routine search and summarisation tasks which in principle can provide creatives, researchers, teachers and content producers time and space to focus on value-adding activities less readily automated. This could help assuage concerns about AI and automation in terms of the potential impact on jobs and labour markets – **AI in publishing has the potential to be creativity enhancing rather than a substitute for human input.** In this way AI technology can be viewed as a ‘productivity enhancer’ for publishing. Other AI solutions can be employed to reduce publishers’ risks (for example, through plagiarism detection) and, further down the value chain, their operational costs (for example, automatically generating invoices, or controlling stock levels) – an important source of potential competitive advantage.

It is important that publishers’ ability to invest in AI is not held back in any way. Our research identified potential investment barriers – some already perceived as significant, some perceived as likely barriers in the near future. It is important that government is aware of these issues and invests in working with industry to ensure these investment benefits are maximised for the UK. The remainder of this section sets out our understanding of these barriers and relevant policy and regulatory issues as a starting point for these discussions.

### 6.1.1 Key policy issues

Our research has revealed three areas of concern or opportunity for publishers’ AI investment activities:

- Providing certainty for investment through a stable intellectual property framework.
- Promoting R&D collaboration between publishers, AI-tech SMEs and academia.
- Helping publishing SMEs access AI investment finance and skills.

### UK Intellectual Property law

While government policy and regulation is not generally viewed as a current AI-investment barrier (by large publishers), 60% of the large industry survey respondents viewed this as at least a potential barrier (Figure 15) and confirmed that the issues of text and data mining (TDM) of copyrighted materials, patenting of AI-generated content, and copyright protection and legal responsibility for AI-generated works are all relevant to publishers' investment decisions (Figure 16).

We recommend that the UK government should ensure that investment certainty is maintained through a stable intellectual property framework. The box below sets out our understanding of the key issues relevant to publishing.

## KEY REGULATORY ISSUES FOR PUBLISHING

Our research has identified the following three key areas:

- **Using TDM to conduct research:** In our industry survey, TDM of copyright protected materials was ranked as the most relevant policy issue for publishers' AI investment decisions (Figure 16, section 4.1.2 ). A key issue is whether and how the legal framework will have to evolve in response to new technologies and products in this area, and the resultant impact this has on investment decisions. We understand that publishers are ready to work with the UK IPO in developing this line of policy thinking.
- **Copyright protection and legal responsibility for AI-generated works:** There are unresolved questions on AI and copyright in many jurisdictions, including the UK, such as: whether subject matter autonomously generated by AI is, or should be, protected by copyright; what degree of human contribution (if any) is required and who the legal owner of the content would be. We heard a range of views on these issues from publishers, including attributing shared responsibility to the individuals and organisations responsible for providing the underlying data and algorithms, and an example earlier in the year of a Chinese court ruling that an AI-generated work was the work of a legal entity<sup>41</sup>. The question of legal responsibility is particularly important for publishers who use AI algorithms to summarise content, or make editorial recommendations where there is scope for AI algorithms to introduce bias or inaccurate representation.
- **Patenting of AI-created content:** Patents are typically awarded to incentivise human creativity. Our research<sup>42</sup> suggests that AI is unlikely to be designated an inventor as invention is considered to require the deployment of human mental facilities. This generates uncertainty over who is the inventor of an invention involving AI activity. One possibility is the human with intellectual responsibility for the AI activity, or alternatively the owner or user of the system.

### R&D collaboration

AI technology is developing at a rapid pace. There is still considerable uncertainty as to which technologies will provide the best fit and deliver the highest benefit. Many publishers (particularly SMEs) do not have the capacity or skills to experiment with and develop AI technologies in-house, and those with some capacity for R&D consider that partnerships or collaborations with specialist AI organisations (for example universities and AI-tech SMEs) offers the most efficient solution to their AI investment needs. Each of our three case studies exemplify collaborative working

<sup>41</sup> <http://ipkitten.blogspot.com/2020/01/another-decision-on-ai-generated-work.html>

<sup>42</sup> Based on the views expressed by panellist Dr Noam Shemtov (Reader in IP & Technology Law, Queen Mary University) during the third session on Ownership, Entitlement and Liability at the June 2019 London conference: 'AI: decoding IP' as summarised in: <http://ipkitten.blogspot.com/2019/06/ai-decoding-ip-conference-day-1.html>



in this way. Publishers that work in this way consider that they add value to these research partnerships through their in-depth knowledge of customer needs and in many cases, their access to data.

Collaborative working between business and academia is at the heart of the UK Government's Industrial Strategy<sup>43</sup>. The Creative Industries Sector Deal<sup>44</sup> is focused on helping creative businesses innovate and access the skills and investment they need. Currently, few publishing companies appear to be benefitting from sector deal or other related programmes and we recommend that this is addressed. One opportunity could be through the Creative Clusters Investment Programme, where publishing businesses located in the programme's cluster areas could benefit from the capacity building activities within the clusters (acquiring knowledge on cutting edge digital technologies) and from the collaboration opportunities with HEI and other creative industry organisations. More broadly, through UK Research and Innovation, a wide range of policies designed to enhance productivity, increase R&D spending and improve collaboration between businesses, academics, the third and public sectors are rolled out under the Industrial Strategy umbrella. These include opportunities for collaborative R&D projects involving full or partial government funding, and for businesses to position themselves in shaping the policy agenda around innovation, including around the use of AI. Given the evidence we have about the potential importance of AI for publishing and the barriers that some businesses in the sector are facing, it is important that publishing has a more active and visible voice participating in and shaping this agenda. Government and industry should work together to identify current and future policies where publishing can play this role.

### **SME access to investment finance and AI skills**

Our industry survey indicates that SME publishers recognise the potential benefits of AI investment but lack the resources they need to invest. The costs of researching, acquiring and implementing AI solutions was considered a significant barrier by over three-quarters of our industry survey SME respondents (Figure 18). AI skills availability was identified as the most significant barrier from a shorter list of barriers which excluded AI costs (see Section 4.2).

Our survey received relatively few responses from SMEs and so we recommend further engagement between the Publishers Association and its SME members. For example, the Publishers Association could use its SME working group (or establish a taskforce focused on addressing SME AI investment issues) to explore these issues further. Evidence from this should inform future engagement with government to establish the best way for publishers SMEs to benefit from the wide range of government programmes focused on improving SME's ability to invest in cutting edge technologies such as AI.

---

<sup>43</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf)  
<sup>44</sup> <https://www.thecreativeindustries.co.uk/media/462717/creative-industries-sector-deal-print.pdf>

# Annex A

## Stakeholder interview guide

### A.1 Interview introduction

Thank you for agreeing to speak to us today to help with our study into the use of Artificial Intelligence in publishing. The overall objective of our study is to report on the role of AI and machine learning in publishing and how these can generate benefits for the sector, consumers and wider society. In this phase we are designing a taxonomy of potential use cases of AI in the publishing supply chain. We have been reviewing the literature on this subject and would like to discuss our thinking to date with you and gather evidence on use cases to start populating our framework (alongside evidence from the literature).

To help guide the conversation we have shared a couple of slides detailing how we are thinking of defining AI applications. At this stage we are considering the academic, education and consumer (fiction and non-fiction books) publishing value chains separately, but we will explore where there are similarities and differences between them in our final analysis. Our value chains include upstream and downstream activities that are not exclusively related to publishing but where opportunities related to AI could still impact the sector.

### A.2 Stakeholder introduction

- Do you have any immediate questions?
- Can we start by you introducing yourself and your organisation
  - Please describe your role in the organisation and any other experience relevant to AI in publishing
  - For your organisation:
    - How many employees, which strand(s), which markets (global?), vertically integrated, suppliers and immediate and final customers?
    - Does the **value chain we've shared with you** chime with your view of the publishing sector?

### Defining AI

First we'd like to consider how you define AI (from a publishing perspective).

Our research suggests that AI is an umbrella term and that Machine Learning sits within it. It also suggests that Natural Language Processing is also a subset of AI, but that only some elements overlap with Machine Learning.

- What is your view please?

Our research has suggested there is not a common framework for defining AI, the various technologies and ways in which it is applied. So our priority has been to develop one for this project. We think it might be useful to define broad 'task areas' where AI is applied. We set this out on our value chain slide (slide 3) in the pack we shared. And categorise the AI that is applied to each part of the value chain in terms of the technology and in terms of the benefits (slide 4)

- Does this approach to categorising use cases of AI in publishing chime with your knowledge and understanding?

- How do you think about AI within your organisation (from a technology perspective, from a benefits perspective, other perspective)?
- **[If relevant]** In your view or experience, does the way AI is talked about in publishing differ from other sectors? How and why?

### **A.3 AI in your organisation today**

We'd like to spend the rest of the conversation walking through the publishing value chain, starting with the parts relevant to your organisation and if time covering upstream and downstream areas too. So for example, if we start at the most relevant dark grey box near to the top of the top left of slide 3, to describe the publishing activity

- Please can you describe use cases of AI, relating this to the framework we set out on slide 4 (if possible)

**[follow conversation through the value chain –explore in each case whether this is common practice, and discuss up-stream and down-stream impacts]**

- Would you describe your organisation as being at the 'cutting edge' in terms of the extent to which it is applying AI? **[explore why]**
- Is your use of AI mostly being driven in-house, or have you worked with others outside publishing to develop this?
- Is this different to how you have used other innovative technologies or methods in the past? How and why?
- Do you see AI being used differently in publishing than in other creative industries? Or other broader sectors? How and why?
- What barriers or blockages have you faced so far for introducing AI in these ways (and could government policy help address any of these)?

### **A.4 AI in your organisation in the future**

**[if not covered in previous conversations]**

- Is your organisation considering other applications of AI?
- Are any other organisations doing this already?
- What barriers or blocks are you encountering/ anticipate for applying AI in these ways (and could government policy help address any of these)?

# Annex B

## The industry survey

This annex sets out details of our industry survey. It provides an overview of the survey design, practicalities of how the survey was administered, the response rate and characteristics of the survey responses.

### B.1 Survey Design

The purpose of our industry survey was to collect primary evidence on the role of AI in publishing. The survey questions were aligned to our taxonomy, which allowed us to both assess and evidence the taxonomy. The survey also covered wider research questions, including the future of AI in publishing and policy relevant concerns.

The survey first asks about the demographics of the respondent, covering size, sub-sector, location and revenues. The survey then covers each of the following topics: defining AI, using and benefiting from AI, the future of AI, barriers to AI use, investment in AI and policy relevant concerns. Figure 24 summarises each section of the survey, and any relevant logic and routing applied.

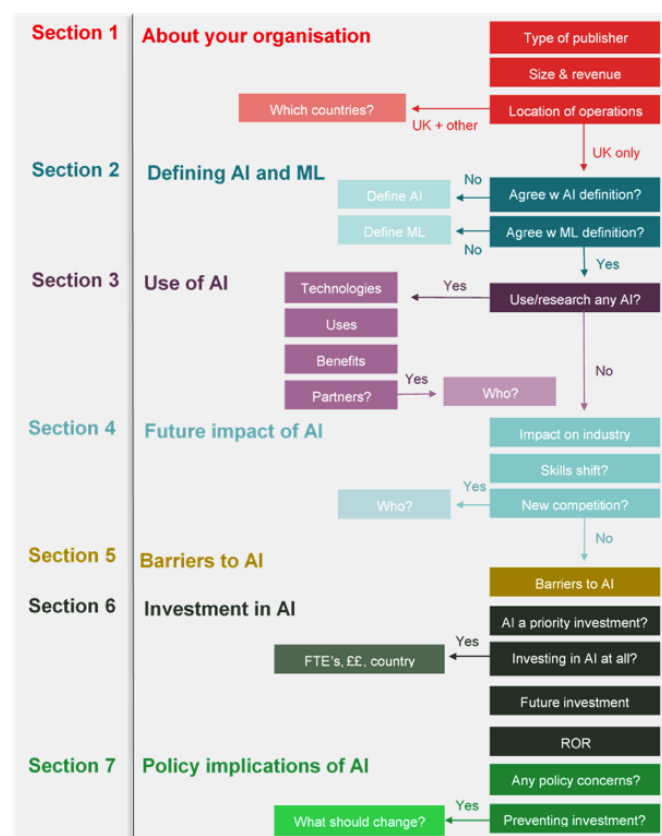
### B.2 Field Work

The survey was distributed online, via the Survey Monkey platform. It was open to PA's 135 members. A link to the survey was emailed to senior level executives in each publisher on 17th February and the survey was live for one month, before closing on the 16th March. Respondents were given the option to leave and return to the survey at any time, as previous responses were saved, giving respondents the option to consult with colleagues<sup>45</sup>. The average completion time was around 10 minutes.

### B.3 Response Rate

The response rate for the survey was 21% of PA's members, with 29 publishers completing at least part of the survey. Of the 29 total responses, 24 of these completed the entire survey (18% of PA's members), indicating a survey completion rate of around 80%. Responding publishers included mainly micro and large publishers, consumer and education publishing, and domestic and global organisations. Figure 25 presents the respondent type by size, sub-sector and where the publisher is headquartered.

Figure 24. Source: Frontier Economics



<sup>45</sup> The exact text shown to respondents was: "When you close the survey, provided you log back in using the same computer and web browser, you will be able to review and edit your earlier responses, enabling you to get information from colleagues if needed."

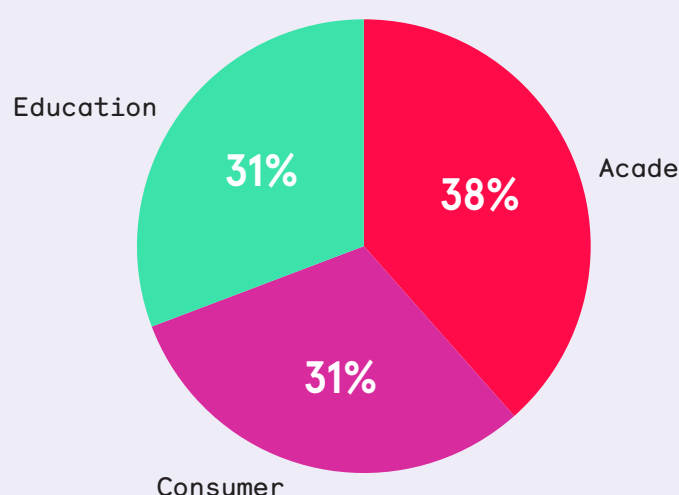
The number of responses to each question (forming the base responses in the results) depends on the following two factors:

- Whether the respondent chose to answer the question. Respondents were more likely to skip questions that were asked towards the end of the survey. This can be seen in which summarises the number of responses and the survey logic. The first question open to all respondents has 29 responses, whereas the final question open to all respondents has 24 responses.
- How the respondent had answered previous questions. The survey was designed to ensure that respondents were only asked questions that were relevant to them, which was inferred from their responses to previous questions.

## B.4 Characteristics of the survey respondents

The survey respondents represent a range of different sizes, sub-sectors and geographical locations.

Figure 25 shows respondent sub-sector, split between education, academic and consumer publishing. Respondents were asked which areas of publishing they were active in and could provide multiple responses. The results show that the three sub-sectors are represented in almost equal proportions in our survey sample. Indeed many respondents consider themselves active in at least two publishing sub-sectors. As a result we have not been able to dis-aggregate results by publishing sub-sector.



**Respondent sub-sector : Which areas of publishing is your organisation active in?**

**Figure 25.**

**Source:** Frontier Economics analysis

**Note:** Base=29 respondents. Question in survey: "Which areas of publishing is your organisation active in?" Respondents were able to select multiple responses and so the results do not infer that 31% of respondents are consumer publishers only for example.

In terms of the size of survey respondents, the majority of respondents are micro (0-9 employees) or large (250+ employees) companies, with 12 and 13 responses respectively.



# Annex C

## Literature review bibliography and sources



Note: weblinks last accessed January 2020

### C.1 Books

1. Beta Writer. 2019. Lithium-Ion Batteries: A Machine Generated Summary of Current Research. <https://rd.springer.com/book/10.1007/978-3-030-16800-1>
2. Duffey, C. 2019. Superhuman Innovation Transforming Businesses Intelligence. Kogan Page. <https://chrisduffey.ai/superhuman-innovation/>
3. Clark, G. and Phillips, A. 2019. Inside Book Publishing. Routledge. <https://routledgetextbooks.com/textbooks/9780415537179/>

### C.2 Reports

4. Cenveo Publishers Services, 2019. AI and NLP in publishing. <https://www.cenveopublisherservices.com/blog/ai-and-nlp-for-publishers>
5. Cenveo Publishers Services, 2017. Six big trends reshaping K12 educational publishing. <https://info.cenveopublisherservices.com/six-big-trends-in-education-publishing>
6. Future TDM. 2016. Baseline report of policies and barriers of TDM in Europe. [https://project.futuretdm.eu/wp-content/uploads/2016/06/FutureTDM\\_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe.pdf](https://project.futuretdm.eu/wp-content/uploads/2016/06/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe.pdf)
7. Future TDM. 2016. European Landscape of TDM Applications Report. [https://project.futuretdm.eu/wp-content/uploads/2016/06/FutureTDM\\_D4.1-European-Landscape-of-TDM-Applications-Report.pdf](https://project.futuretdm.eu/wp-content/uploads/2016/06/FutureTDM_D4.1-European-Landscape-of-TDM-Applications-Report.pdf)
8. Future TDM. 2016. Research Report on TDM Landscape in Europe. <https://project.futuretdm.eu/wp-content/uploads/2016/05/D3.1-Research-Report-on-TDM-Landscape-in-Europe-FutureTDM.pdf>
9. Luckin, R., Holmes, W. 2017. Intelligence Unleashed: An argument for AI in Education. <https://discovery.ucl.ac.uk/id/eprint/1475756/>
10. SpotOn/Biomed. 2017. What might peer review look like in 2017? [https://figshare.com/articles/What\\_might\\_peer\\_review\\_look\\_like\\_in\\_2030\\_/4884878](https://figshare.com/articles/What_might_peer_review_look_like_in_2030_/4884878)

### C.3 Papers

11. Spangler, S. et al. 2014. Automated Hypothesis Generation Based on Mining Scientific Literature. <https://scholar.harvard.edu/files/alacoste/files/p1877-spangler.pdf>
12. McKinsey Global Institute, 2018. Notes from the AI frontier – Insights from hundreds of cases. <https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20ai%20frontier%20applications%20and%20value%20of%20deep%20learning/notes-from-the-ai-frontier-insights-from-hundreds-of-use-cases->

## C.4 Surveys

13. Econsultancy, 2018. Digital Trends 2018. <https://www.adobe.com/content/dam/acom/en/modal-offers/pdfs/0060629.en.aec.whitepaper.econsultancy-2018-digital-trends-US.pdf>
14. Gould Finch, 2019. The future impact of artificial intelligence in the publishing sector. <https://gouldfinch.com/en/the-future-impact-of-artificial-intelligence-on-the-publishing-industry/>
15. Pfeiffer, A., 2018. Creativity and technology in the age of AI. <https://www.pfeifferreport.com/essays/creativity-and-technology-in-the-age-of-ai/>
16. Teradata, 2017. State of artificial intelligence for enterprises. [https://assets.teradata.com/resourceCenter/downloads/ExecutiveBriefs/EB9867\\_State\\_of\\_Artificial\\_Intelligence\\_for\\_the\\_Enterprises.pdf](https://assets.teradata.com/resourceCenter/downloads/ExecutiveBriefs/EB9867_State_of_Artificial_Intelligence_for_the_Enterprises.pdf)
17. UNSILO. 2019. UNSILO AI in Academic Publishing Survey. <https://unsilo.ai/wp-content/uploads/2019/11/UNSILO-Survey-on-AI-in-Academic-Publishing-2019.pdf>

## C.5 Electronic articles

18. Blockdelta, 2019. 10 artificial intelligence technologies that will dominate in 2019. <https://blockdelta.io/10-artificial-intelligence-technologies-that-will-dominate-in-2019/>
19. Business Wire, 2017. Yewno launches unique new AI tool. <https://www.businesswire.com/news/home/20170530005384/en/Yewno-Launches-Unearth-Unique-New-AI-Tool>
20. Chen, H. 2019. The impact of AI on peer review. <https://medium.com/digital-publishing-strategy/the-impact-of-ai-on-peer-review-628f1318b18c>
21. Digital Trends, 2016. Japanese AI wrote a novel almost wins a literary award. <https://www.digitaltrends.com/cool-tech/japanese-ai-writes-novel-passes-first-round-national-literary-prize/>
22. Forbes, 2019. Artificial intelligence can now write amazing content what does this mean for humans? <https://www.forbes.com/sites/bernardmarr/2019/03/29/artificial-intelligence-can-now-write-amazing-content-what-does-that-mean-for-humans/#41e3a65550ab>
23. McKinsey & Company, 2018. Mapping AI techniques to problem types. <https://www.mckinsey.com/industries/financial-services/our-insights/mapping-ai-techniques-to-problem-types>
24. McKinsey & Company, 2018. What AI can and can't do (yet) for your business <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/what-ai-can-and-cant-do-yet-for-your-business>
25. Penn, J. 2019. 9 ways that Artificial Intelligence will disrupt authors and the publishing industry. <https://www.thecreativepenn.com/2019/07/01/ai-disruption-publishing-authors/>
26. Price, S., Flach, P. 2017. Computational support for academic peer review: a perspective from Artificial Intelligence. <https://m-cacm.acm.org/magazines/2017/3/213825-computational-support-for-academic-peer-review/fulltext?mobile=true?mobile=true>
27. Times Higher Education, 2016. Robot written reviews fool academics. <https://www.timeshighereducation.com/news/robot-written-reviews-fool-academics>
28. Upshall, M. 2019. Using AI in scholarly publishing. <https://insights.uksg.org/articles/10.1629/uksg.460/>