



Department for  
Science, Innovation  
& Technology



THE  
BEHAVIOURAL  
INSIGHTS  
TEAM

# AI-Assisted vs human-only evidence review

Results from a comparative study

Dr. Mark Egan, Lauren Leak-Smith, Antonio Hanna-Amodio, Maxime Sirera

April 2025

**Note:**

This research was supported by the R&D Science and Analysis Programme at the Department for Culture, Media & Sport. It was developed and produced according to the Behavioural Insights Team's research team's hypotheses and methods. Any primary research or findings do not represent Government views or policy. Please note that this research was commissioned under the previous government of 11 May 2010 to 5 July 2024, and before the founding of the UK Metascience Unit.



© Crown copyright 2025

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk).

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to the Strategic Evidence Team at: [set\\_mailbox@dsit.gov.uk](mailto:set_mailbox@dsit.gov.uk)

# Contents

Executive Summary	4
Background, Research Topic and Methodology	5
Background	5
Research Topic	6
Methodology	7
Key Results: How the two reviews compared	9
Key Finding 1: Speed	9
Key Finding 2: Quality – Reference list	10
Finding 4: Mechanisms	13
Key Finding 5: Conclusions	15
Additional Findings	16
BIT's four step-process for implementing AI	16
When AI goes wrong	21
Reflections from the producers and consumers of the reviews	23
Reflections from the human and the AI-assisted researcher	23
Reflections from the QA process	24
A final test	25
Discussion and Limitations	26
What we think these results mean	26
Limitations of our approach	27
Appendix	28
Appendix 1: List of references	28
Appendix 2: AI-led prompt	33

# Executive Summary

BIT ran a comparative exercise in 2024 with the UK Department for Culture, Media & Sport (DCMS) and Department for Science, Innovation & Technology (DSIT) to investigate the robustness and reliability of **using generative AI to help produce rapid evidence reviews**.

Two BIT researchers separately conducted reviews on the topic “How technology diffusion impacts UK growth and productivity”. Both got the same briefing and inclusion/exclusion criteria, but one review was produced by ‘human-only’, the other was ‘AI-assisted’ using a mix of tools (ChatGPT 4, Claude 2, Elicit & Consensus).

The **AI-assisted output was ultimately completed in 23% less time** – it particularly excelled at speeding up the process of analysing and synthesizing studies. But, the initial draft of the AI output was also judged to be somewhat stilted; it therefore required more revisions than the ‘human’ version.

The two finalised reviews were ultimately similar in quality. They both produced (i) credible, non-identical reference lists of ~20 studies each; (ii) 6 evidence-based mechanisms through which new technology impacts growth and productivity, of which 4 were thematically similar; and (iii) 3 conclusions, of which 2 were thematically similar.

We recognise that this study is effectively a case-study, and that the results are not generalisable. The ability of both humans and AI models to review literature will vary substantially, including by topic. However, we think AI has the potential to enhance the process of conducting rapid evidence reviews. It is not yet a game-changer – it still produces occasional, peculiar hallucinations and errors which mean its outputs require manual verification. However, AI is improving quickly, so these issues may soon be reduced.

**We therefore recommend that more work be undertaken to understand how and when AI can be implemented in evidence reviews.** In this case study, Large Language Model (LLM) tools were found, on this occasion, to have sped up the process of analysing selected literature – for this phase of the literature review, the AI-assisted process took 56% less time. They also proved effective in synthesising credible overall summaries. However, it is important that researchers take the time needed to learn to use these AI tools effectively: precise, detailed and explicit prompts were found to impact these tools’ efficacy. Further research is needed to clarify the benefits and limitations of the technology.

Phase	Human	AI-assisted
Total number of hours	117.75	90.5

Scanning	23	16
Selection	10	14
Analysis	34	15
Synthesis	32.5	18.5
Revisions	18.25	27

# Background, Research Topic and Methodology

## Background

*We ran a comparative trial to test whether AI tools could improve the process of conducting evidence reviews.*

Generative AI tools, including large language models like ChatGPT, have the potential to boost productivity across various sectors by automating routine tasks, enhancing human creativity and providing instant access to vast amounts of information.

One emerging use-case is the application of Generative AI to the process of conducting evidence reviews - studies which collate and examine the best available academic evidence on a particular topic. Conducting these reviews typically involves processing, analysing and synthesizing vast amounts of text data – a process which seems particularly suited to the strengths of AI. Evidence reviews are common in government, academia and industry, but can be laborious and time-consuming to conduct, often requiring the manual identification and analysis of dozens or hundreds of research studies.

Despite the promise of AI for speeding up this process, the UK government guidance notes that “output from generative AI is susceptible to bias and misinformation”. The known tendency of the technology to also ‘[hallucinate](#)’ false facts means it is currently unclear whether its application would indeed translate into genuine efficiency gains (i.e. a speeding-up of the process of conducting reviews without unduly compromising the robustness of the output).

To investigate the robustness and reliability of evidence reviews produced using Generative AI, the Behavioural Insights Team (BIT) ran a comparative exercise over January to March 2024 in partnership with the Department for Culture, Media & Sport (DCMS) R&D programme & Department for Science, Innovation & Technology (DSIT) strategic evidence team. Specifically, we conducted two rapid evidence reviews on the same topic: one was produced by a ‘human-only’, the other was ‘AI-assisted’.

This report describes the method and results of that exercise.

## Research Topic

*The topic of the reviews was “How technology diffusion impacts UK growth and productivity”.*

Productivity was conceptualised as outputs relative to resource used; growth as an increase in value over time.

The reviews were ‘rapid’ rather than ‘comprehensive’, meaning they prioritised covering the major pieces of relevant research within a few weeks, rather than taking months to systematically document all research on the topic – although they did include both academic studies and ‘grey’ literature (e.g. government reports). The specific technologies being examined were informed by DSIT’s [Science and Technology Framework](#), which identifies AI, engineering biology, future telecommunications, semiconductors, and quantum technologies as key to the UK’s strategic advantage and economic growth.

The reviews focused on identifying (i) the ‘how’ rather than the ‘what’ – clarifying the mechanisms through which the diffusion of new technologies could improve growth/productivity in a UK context, rather than merely documenting their impact, and (ii) the practical takeaways of these findings.

To help sharpen the focus of the potentially broad and complicated topic under review, BIT and DCMS/DSIT agreed to include studies which (i) examined the UK (or areas with similar socioeconomic characteristics such as USA, France, Germany), (ii) were published since the year 2000, (iii) drew more on empirical evidence rather than theory alone, (iv) documented impact at the business level, (v) examined technological changes taking place recently (late 20th century / 21st century).

The reviews excluded (i) non-English-language material and (ii) studies or history texts examining multi-faceted technological / societal shifts rather than discrete technology changes.

## Methodology

*To isolate the effect of the AI tools, the process of running the two reviews was standardised as much as possible.*

We ran two evidence reviews on the same topic: one was conducted using a 'human only' approach, the other had access to the latest AI tools.

In an attempt to cleanly measure the impact of AI, we aimed to standardise the process of conducting the reviews as much as possible, so that only notable difference between them would, in theory, be the use of AI tools.

The reviews were therefore conducted in parallel by two junior researchers from BIT, who worked apart from each other to avoid contamination. Both staff members were in the same type of role, had been at BIT for a similar amount of time (~2 years) and had previous experience producing evidence reviews. Both received the same written briefing about the goal and scope of the review, used the same high-level search terms, were given the same inclusion/exclusion criteria to apply to the studies, received the same number of regular briefings about the timelines and milestones of the project, were provided the same amount of ring-fenced time to conduct the task (but could use more or less if required), and got the same detailed templates for writing-up their results. Both also had their work and outputs monitored and reviewed by experienced senior BIT staff members.

The evidence reviews were assessed on two high-level dimensions: speed and quality.

Speed was measured by comparing the hours recorded by each researcher to different phases of the project. Quality was measured by assessing the extent to which the reviews captured an appropriate set of key references, drew accurate and relevant insights from these studies, and synthesized them into useful conclusions. We also monitored how effectively the two reviews passed the 'market-test', by recording the feedback provided by our partners at DCMS & DSIT, the ultimate customers of the reviews, on the initial drafts.

*Our initial plan for how the 'Human' vs 'AI-assisted' approaches would differ at each phase of the review.*

The table shows our initial plan for how the two review approaches would differ.

The BIT project team were very experienced with the process of running 'human' evidence reviews (i.e. without AI tools). We expected that the process of conducting that review would go according to plan - this proved to be the case.

However, although we had a theoretical understanding of how AI might be applied to the four stages of the review process, the researcher conducting that review was also encouraged to 'learn-by-doing' and use their initiative to apply the tools in whatever manner they deemed most effective.

Phase	Human approach	AI-assisted approach
1. Scanning	Manually enter search terms into Google + Google Scholar, collate candidate papers; apply 'snowball' methodology to identify further candidates.	Use AI to suggest papers; Alter prompts to find more papers on AI tools.
2. Selection	Manually screen out candidate papers based on inclusion/exclusion criteria.	Provide AI with list of links to academic papers + ask it to flag which should be included / excluded based on criteria; Supplement with manual check.
3. Analysis	Manually review and summarise papers.	Use AI to produce a high-level summary of each paper, then manually edit.
4. Synthesis	Manually write Executive Summary concisely communicating key takeaways.	Use AI to produce an overall summary / key takeaways, then manually edit.

*A mix of AI tools were used for the AI-assisted review.*

Ultimately, we used a mix of AI tools depending on the phase of the evidence review.

This was because no single AI tool currently exists that can effectively conduct every step of the literature review process.

Following personal experimentation aimed at identifying which tools work best for what task (e.g finding literature or summarising papers), we focused on a final set of 4 AI tools.

More detail on the implementation aspect of the tools, as well as the tools we deemed less relevant for this task, can be found in [Additional Findings I](#).

Phase	AI tool used	Rationale
1. Scanning	<a href="#">Elicit</a> and <a href="#">Consensus</a>  Free versions	Both tools focus on finding scientific papers based on inputting a prompt as a research question.  We used both tools as Elicit shares comparatively more papers, while Consensus currently limits



		results to a smaller sample of key papers. On occasion, Consensus found papers Elicit missed.
2. Selection	<a href="#">Claude 2 (Pro)</a> Paid version	Large PDFs can be uploaded to Claude and, as such, it can make an assessment on which papers to include if you also share your inclusion/exclusion criteria.
3. Analysis	<a href="#">Claude 2 (Pro)</a> Paid version	Claude can quickly and concisely summarise key areas of interest from key papers, such as methodological details and the sample studied.
4. Synthesis	<a href="#">Claude 2 (Pro)</a> and <a href="#">ChatGPT 4</a> Paid versions	ChatGPT can help with a wide range of writing tasks: from general report drafting to editing and improving clarity of more specific portions of text.

## Key Results: How the two reviews compared

### Key Finding I: Speed

*The AI-assisted review was completed in 23% less time than the human review (~90 vs ~118 hours).*

Phase	Human	AI-assisted	Comments
Total # hours	117.75	90.5	AI in 23% less time overall.
Scanning	23	16*	AI in 30% less time. When searching for papers, AI creates paper summaries which makes it easier to assess their relevance. The human researcher often needed to open the paper and scan it to ascertain whether it should be put on the longlist for potential inclusion. *Includes 4 hours spent by the AI-assisted researcher assessing which AI tools to use.

Selection	10	14	Human in 29% less time. Having engaged more deeply with the papers during the scanning phase, the human researcher was able to more quickly select which studies to include in the final sample.
Analysis	34	15	AI in 56% less time. AI excels at quickly summarising & analysing papers, complemented by checks and probing from the researcher. This phase was particularly time-consuming for the human researcher because they needed to manually scrutinise many papers which did not explicitly discuss the mechanisms underpinning their results - a key focus of our analysis.
Synthesis	32.5	18.5	AI in 43% less time. Synthesising ~20 detailed technical papers required a lot of careful reading and note-taking for the human researcher; AI does not need to 'warm up' in the same way; it produced credible summaries very quickly, which could then be further iterated and interrogated.
Revisions	18.25	27	Human in 32% less time. It took longer to produce a draft of the human report, but our DCMS/DSIT partners assessed it as stronger than the AI-assisted first draft. It consequently required less time for revisions.

## Key Finding II: Quality – Reference list

*Both reviews produced credible reference lists.*

An effective evidence review must identify the right studies – ones relevant to the research question and robust enough to provide generalisable insights.

The scanning phase, which involved high-level searches and assessments of study-relevance, produced longlists of 77 studies in the 'Human' review, and 35 studies in the 'AI review'. This difference may have been partly due to AI tools summarising key paper details by default, making it easier to judge if they were relevant at first glance.

The selection phase, which involved closer scrutiny of the candidate papers and rigorous application of the inclusion/exclusion criteria, reduced these to shortlists of 20 studies in the Human review, and 22 studies in the AI review. Both shortlists focused on papers identifying mechanisms through which new technologies could impact growth and productivity, but also included some related to technology 'adoption' specifically.

Details of both shortlists are in the [Appendix](#). As implied by the adjacent tables, the two shortlists were not identical. However, our assessment was that both shortlists were credible: they drew from reputable sources, included papers by well-known researchers in the area, and had article titles clearly relevant to the research question.

**Table 1: Example studies in both reviews:**

Ballestar, María Teresa, Ángel Díaz-Chao, Jorge Sainz, and Joan Torrent-Sellens. "Knowledge, robots and productivity in SMEs: Explaining the second digital wave." <i>Journal of Business Research</i> 108 (2020): 119-131.
Czarnitzki, Dirk, Gastón P. Fernández, and Christian Rammer. "Artificial intelligence and firm-level productivity." <i>Journal of Economic Behavior &amp; Organization</i> 211 (2023): 188-205.
DeStefano, Timothy, Richard Kneller, and Jonathan Timmis. "Cloud computing and firm growth." <i>Review of Economics and Statistics</i> (2023): 1-47.

**Table 2: Example studies only in:**

Human review	AI-assisted review
Brynjolfsson, Erik, Wang Jin, and Kristina McElheran. "The power of prediction: predictive analytics, workplace complements, and business performance." <i>Business Economics</i> 56 (2021): 217-239	Bartel, Ann, Casey Ichniowski, and Kathryn Shaw. "How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills." <i>The Quarterly Journal of Economics</i> 122, no. 4 (2007): 1721-1758.
Dixon, Jay, Bryan Hong, and Lynn Wu. "The robot revolution: Managerial and employment consequences for firms." <i>Management Science</i> 67, no. 9 (2021): 5586-5605.	Crespi, Gustavo, Chiara Criscuolo, and Jonathan Haskel. "Information technology, organisational change and productivity." (2007).
Stucki, Tobias. "Which firms benefit from investments in green energy technologies? - The effect of energy costs." <i>Research Policy</i> 48, no. 3 (2019): 546-555.	Graetz, Georg, and Guy Michaels. "Robots at work." <i>Review of Economics and Statistics</i> 100, no. 5 (2018): 753-768.

*There was surprisingly little overlap in the two lists of references.*

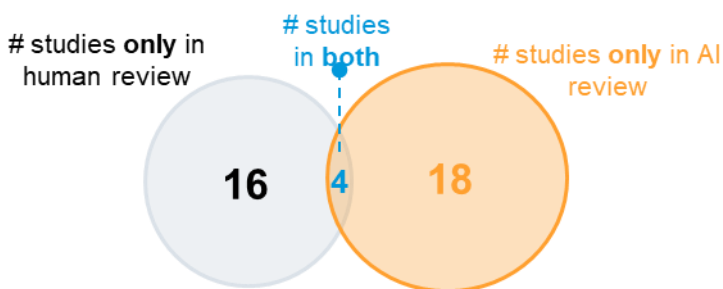
There were 16 references only in the human-only evidence review, 18 only in the AI-assisted evidence review and 4 in both.

We think the low overlap between the two reference lists was influenced by two important factors:

(i) AI-driven differences: The models that dictate how Elicit and Consensus rank and generate papers differs to Google Scholar. As a result, they return different results even when searching for the same research question.

(ii) Researcher-level differences: The two researchers were given the same briefing and inclusion/exclusion criteria, but there remained an element of subjective judgement in deciding which papers were most important / relevant.

Both factors likely impacted which papers the two researchers ended up including in their review. As alluded to previously, the divergence was likely also driven by the complexity of the topic, and the limited time available with which to search through it: if the scanning portion of the reviews had been conducted for several months, rather than 2 weeks, the two lists would likely have expanded and overlapped much more.



**Figure 1: Overlap of studies in the human-only and AI-assisted evidence reviews**

Even when using the same search (“What is the impact of technology diffusions on growth and productivity in the UK?”), there was no overlap of the papers on the first page of Google Scholar vs Elicit.

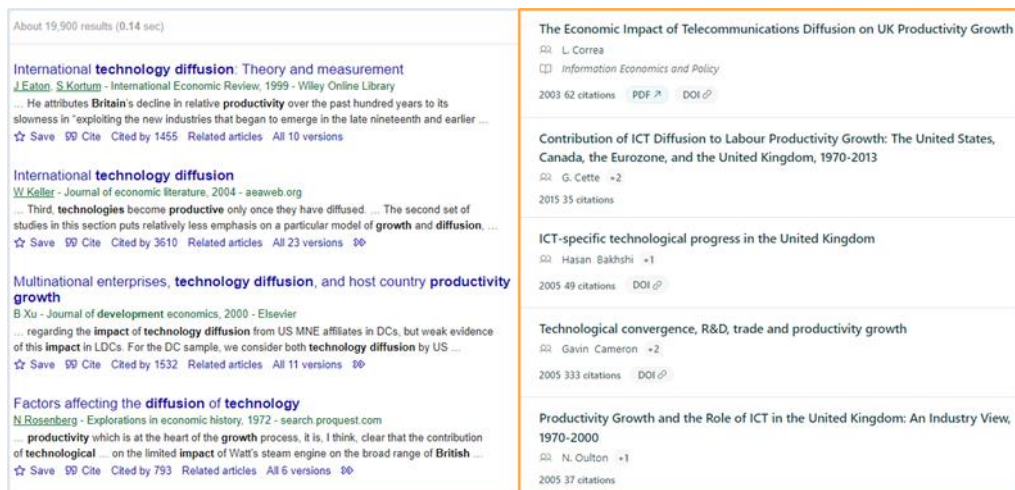


Figure 2: Comparison of Google Scholar and Elicit search results

## Finding 4: Mechanisms

Both studies identified 6 evidence-based mechanisms through which the diffusion of new technology impacts growth and productivity, of which 4 were thematically similar

	Human	AI-assisted	Common theme
1	Training, management expertise & upgrading organisational processes are crucial to unlock the benefits of new technologies.	Technology diffusion enhances businesses' organisational and operational capabilities, impacting productivity and growth.	Complementary investments
2	Firms with more in-house expertise and human capital reap more benefits from new technologies.	Broad knowledge gains flow from technology adoption and boost productivity.	Human capital
3	Infrastructure & regulatory environments influence the likelihood of adopting technologies.	Trade openness encourages technology adoption, improving productivity.	Enabling macro-environment
4	Larger firms with access to more resources are more able to absorb the high costs of technological investments.	New technology can lower fixed costs and barriers to entry into markets, particularly for smaller firms.	Cost barriers

5	Young firms that aren't already reliant on existing technologies are more likely to capitalise on the arrival of new ones.	Advances in robotics & manufacturing technologies enable automation and improve production efficiency.	-
6	After adopting new technology, a period of learning-by-doing may be needed to unlock their benefits.	New general-purpose technologies, such as AI, spur follow-on innovation throughout firms.	-

## Key Finding 5: Conclusions

*Both studies produced 3 conclusions, of which 2 were thematically similar*

	Human	AI-assisted	Common theme
1	Financial incentives and support could help businesses adopt and commercialise new technology, especially smaller firms. SMEs may require additional resources to overcome initial cost barriers and facilitate their adoption of productivity-enhancing technologies, for example by optimising the design of existing support measures, such as R&D tax relief, through timely prompts and targeted campaigns	Access to and ‘early adoption’ of general-purpose technologies is crucial. The benefits of early technological adoption can persist, and can be more pronounced for smaller firms. A timely new technology is AI - fostering environments that support its diffusion across UK business and sectors should ensure both equitable access to these technologies as well as productivity gains.	Better technology adoption for smaller firms especially
2	Testing and evaluating ways to improve take-up can help upgrade existing business management and tech support schemes. The UK government’s ‘Help to Grow - Management’ programme has gotten low take-up. Greater use of peer-to-peer networking may encourage take-up of this programme; more broadly business support schemes may need to be tailored depending on firm size, sector and age.	Addressing local skill gaps can support less productive firms in adopting new technologies. Firms often become more productive as they adopt new technologies. However, productivity increases are only achieved if employees have the right skills to effectively use new technologies. Ensuring that any skill gaps are addressed is therefore essential. Managers can play an important role by creating an environment that encourages innovation and supporting learning among employees.	Improved managerial practices and workforce skills
3	De-shrouding B2B markets can improve the quality of complementary investments. Complementary investments (e.g. training, IT infrastructure upgrades)	In manufacturing contexts, firms should embrace advances in robotics for task automation. There are clear and positive impacts of technological adoption in manufacturing on productivity,	-

	are essential to unlock the full value of new technologies.	efficiency, and economic competitiveness.	
--	---	---	--

# Additional Findings

## Additional Findings I: BIT’s four step-process for implementing AI

*The four-step process for conducting the AI-assisted review*

**One:** Enter a research prompt into Elicit. We started with Elicit as it outputs comparatively more literature. Other useful features include it summarising papers’ abstract by default and further filtering options, such as by outcome measure and main findings.

**Two:** We then complemented this approach by inputting the same prompt into Consensus. This was primarily to check if we had missed relevant literature. Consensus currently aims to provide concise one line summaries for up to ten research papers and also includes other features, such as indicating how influential the journal the paper is submitted in is.

**Three:** Screening and summarise papers was done using Claude. We uploaded PDFs of the papers identified during steps 1 and 2 into Claude, alongside our inclusion criteria. We prompted it to screen the paper against our criteria, provide a yes/no decision on whether it should pass screening (alongside a longer written justification) and to summarise the paper generally. Notably, we quickly fact checked its results were consistent with the paper and the researcher made the ultimate judgment as to whether it would be included in the final review.

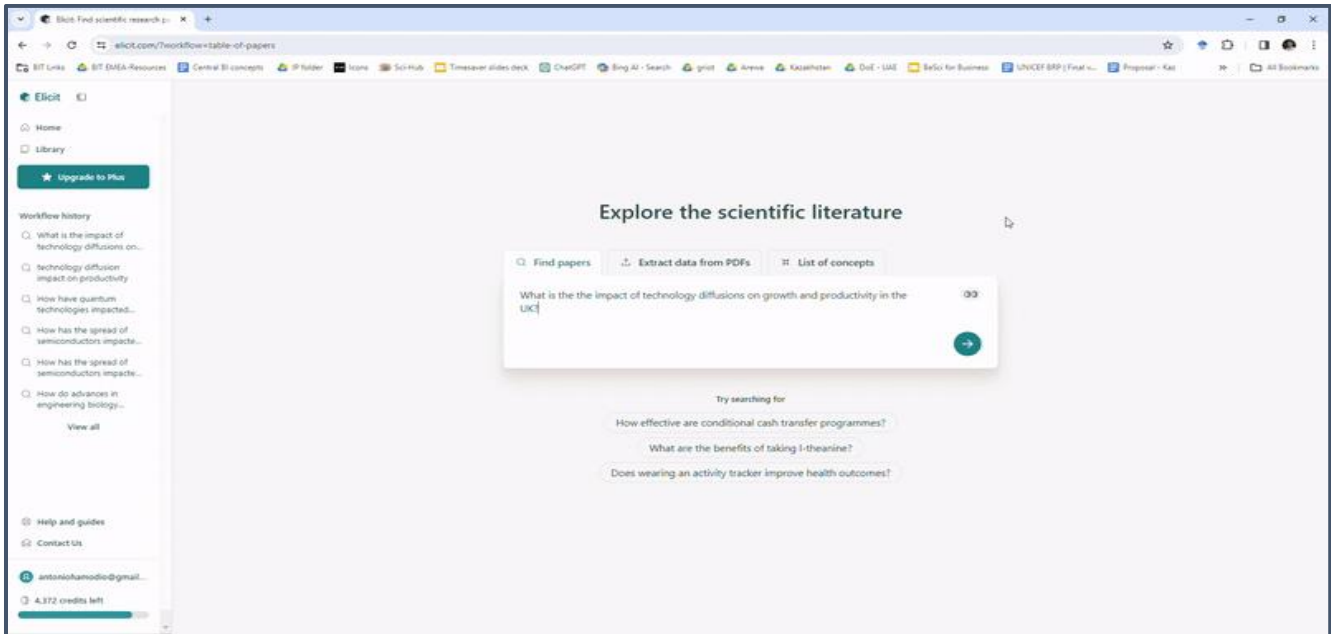
**Four:** For the report writing itself, we used both Claude and ChatGPT as general-purpose writing aids. For example, by uploading our spreadsheets containing paper summaries, key paper details and our internal notes into Claude and asking it to write a first draft for a mechanism section. Or feeding ChatGPT relevant information to speed up the drafting of the more straightforward and formulaic sections of the report (e.g intro and methodology sections).

*Step 1: Enter prompts into Elicit as specific research questions, then click through to individual papers.*

- We entered specific research questions into Elicit
- It then gave us a paragraph summary of the first 4 papers found, as well as a table of more papers



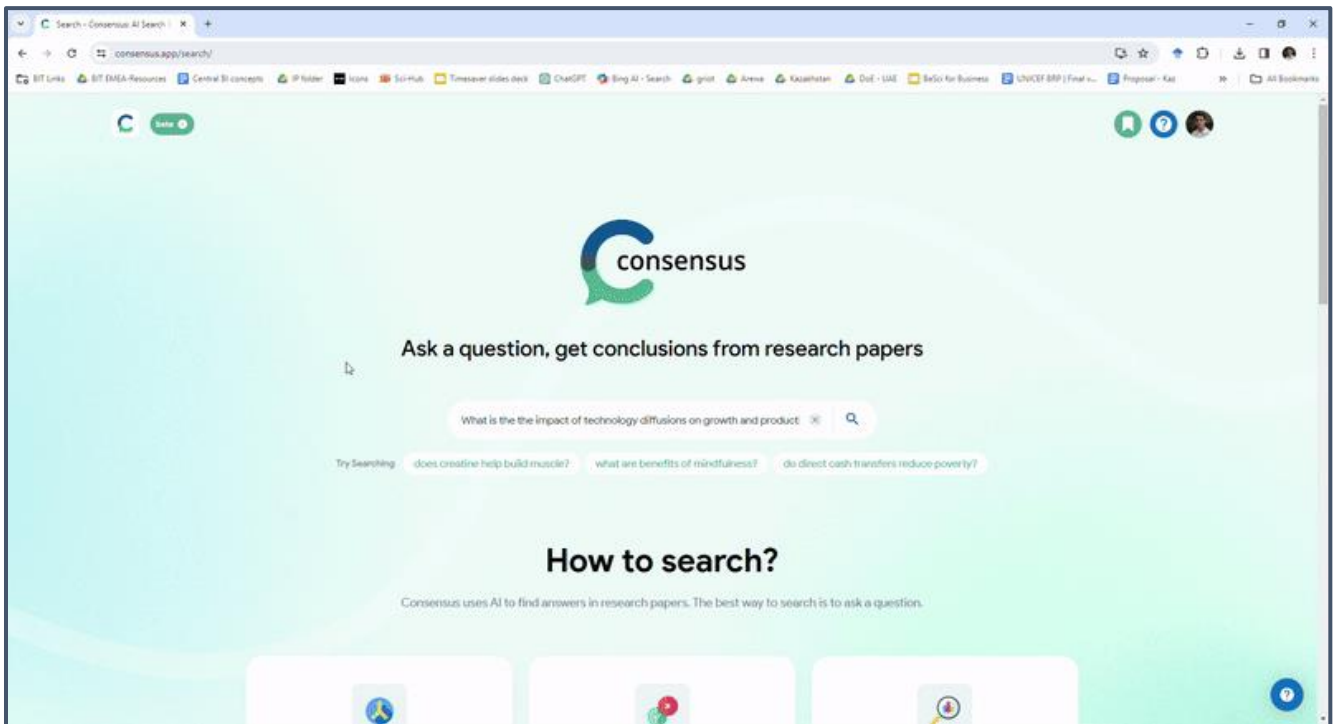
- Based on the paper titles and abstract summaries in the table, we decided whether to look into them further
- We then assessed the full paper text on Elicit or downloaded the PDF directly, if available



**Figure 3: The process of using Elicit to find research papers**

*Step 2: Repeat the process on Consensus to see if we missed any key studies and consider any new papers.*

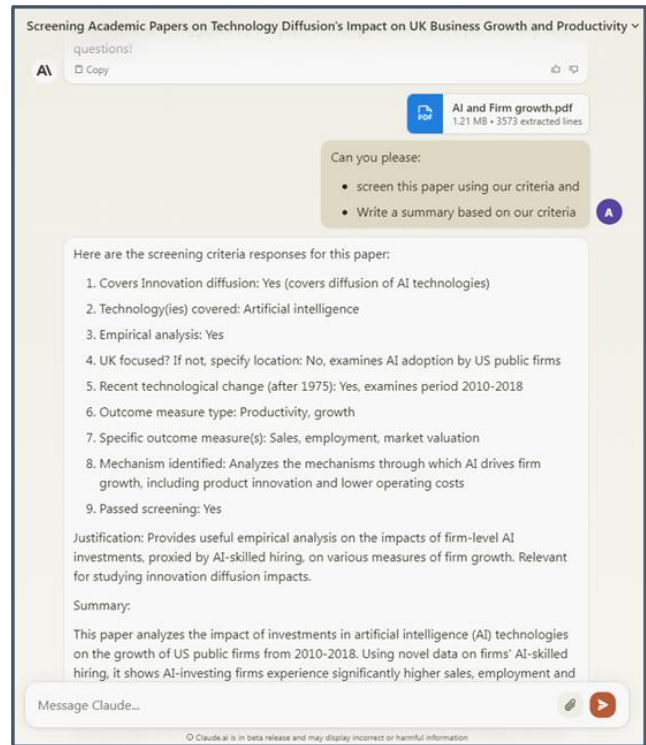
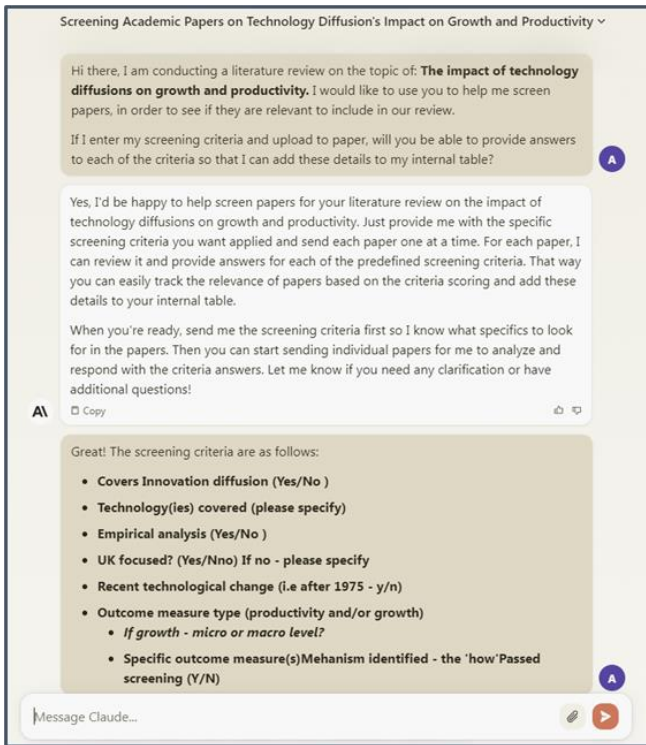
- We entered the same research questions into Consensus
- It then gave us a list of the 10 most relevant papers it found, along with a one sentence key takeaway for each paper
- Based on the paper titles and key takeaways, we decided whether to look into them further
- We then clicked through and were redirected to journal pages, or downloaded the PDF directly, if available



**Figure 4: The process of using Consensus to find research papers**

*Step 3: Use Claude to screen papers*

- This is the initial prompt we used
- We then fed it our specific inclusion criteria, to use for all papers we would upload
- Next, we uploaded the paper and instructed the AI tool to screen it according to the criteria we had previously shared, as well as to summarise the paper more generally



**Figure 5: Using Claude for screening**

*Step 4: Use Claude and ChatGPT to analyse and synthesise studies, and assist with writing in general.*

Here is an example of prompt used to produce a first draft of one of the sections outlining a key mechanism. Note that we identified the specific papers that the model should base its draft on. Earlier in the chat, we also inputted the specific format we wanted for each paragraph (see Appendix 2).

Let's move onto the next mechanism.

**Mechanism 4: Automation and boosting production efficiency**

The papers I want you to base your write-up on are:

1. **Graetz, Georg, and Guy Michaels. "Robots at work." Review of Economics and Statistics 100, no. 5 (2018): 753-768.** Ballestar, María Teresa, Ángel Díaz-Chao, Jorge Sainz, and Joan Torrent-Sellens. "Knowledge, robots and productivity in SMEs: Explaining the second digital wave." *Journal of Business Research* 108 (2020): 119-131.
2. **Bartel, Ann, Casey Ichniowski, and Kathryn Shaw. "How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills." The quarterly journal of Economics 122, no. 4 (2007): 1721-1758.**
3. **Ballestar, María Teresa, Ángel Díaz-Chao, Jorge Sainz, and Joan Torrent-Sellens. "Impact of robotics on manufacturing: A longitudinal machine learning perspective." Technological Forecasting and Social Change 162 (2021): 120348.**
4. **Ballestar, María Teresa, Ángel Díaz-Chao, Jorge Sainz, and Joan Torrent-Sellens. "Impact of robotics on manufacturing: A longitudinal machine learning perspective." Technological Forecasting and Social Change 162 (2021): 120348.**

Remember to points raised above, such as mention the country/countries that the study looked at, the technologies, the methodologies they used, and relative strengths and weakness of the study.

Do you have any questions before you begin?

Write-up: AI report

Here is a draft summary of the evidence on automation and boosting production efficiency as a mechanism linking technology diffusion to productivity/growth:

A 2018 econometric analysis of industrial robotics adoption in 14 industries across 17 developed countries including the UK over 1993-2007 leverages instrumental variables to establish causal impacts. The panel data regression finds robots increased labor productivity, wages and total factor productivity while lowering output prices - but did not reduce total employment significantly. However, robot densification did diminish the share of low-skilled employment, indicating potential worker displacement. Reliance on proxy technology measure and incomplete productivity metrics pose limitations.

A 2020 Spanish study on industrial robotics adoption by manufacturing SMEs over 2008-2015 combines survey data and multiple productivity indicators in a multivariate analysis. It determines robots associate with superior financial performance, productivity and higher wages - but knowledge spillovers that initially complement these automation gains fade over time. However, tech savvy large firms show additional lasting benefits unlike smaller enterprises. The single country basis and reliance on self-reported technology use temper wider applicability.

A 2021 machine learning analysis of nearly 4600 Spanish manufacturers over 1991-2014 uncovers that robot-adopting firms display significantly better economic results than non-adopters. It also finds both small and large companies benefit, though initial productivity gains plateau faster for larger enterprises as they mature in utilizing the technology. While highlighting the promise of smarter statistical techniques, limited years observed after adoption somewhat constrain insights into lasting impacts.

A 2007 study combining proprietary plant-level operations data in the US valve manufacturing sector with site visits and interviews determines that advanced IT equipment (like robots) boosts productivity while enabling greater product customization. The reduced set-up times and operating costs due to substituting automated processes for manual ones increase performance and custom offerings. But this qualitative method does

Reply to Claude...

Figure 6: Using Claude to explain a section outlining a key mechanism

Here is an example of the prompt we used to create an initial draft for the introduction of the report. As we note later in the report, it was important to be as specific and detailed as possible with prompts to get higher quality outputs.

**You**

Perfect - here is what should be covered in the introduction:

Introduction (1-2 pages)

1. Overview of the key technology innovations of the late 20th and 21st century, and their impact on growth and productivity.

- my thoughts: here, after general outline of the key technological developments, can quote some of the studies that focus more on effect sizes + add some colour + mention Solow's paradox: The concept is attributed to Robert Solow, in reference to his 1987 quip, "You can see the computer age everywhere but in the productivity statistics."<sup>[2]</sup>

- Finish this bit with the fact there are important questions about the relative differences in impact across different technologies in varying contexts - also make point that time cannot be sped up, so the effect longer-term impacts of technologies can only be and should be continually evaluated. This links to next main section:

2. Purpose of the review & research question.

- my thoughts: say that the purpose of this rapid evidence review is to provide a practical summary of the empirical evidence to date, with a particular UK focus →  
 - mention the UK's 'productivity problem' - compared to other similar countries

Before you start, do you have any clarifying questions?

**ChatGPT**

No clarifying questions at the moment. Based on your inputs, let's refine the introduction to incorporate these specific elements:

**Introduction**

The late 20th and early 21st centuries have been marked by unprecedented technological advancements, fundamentally altering the landscape of global economies and their productivity patterns. Innovations such as the internet, mobile computing, artificial intelligence, and renewable energy technologies have not only revolutionized the way we live and work but have also played a pivotal role in driving economic growth and enhancing productivity. Despite the rapid adoption and visible impact of these technologies, quantifying their effect on productivity has presented challenges, famously encapsulated in Solow's Paradox. This phenomenon, highlighted by economist Robert Solow's 1987 observation, "You can see the computer age everywhere but in the productivity statistics," underscores the complex relationship between technological innovation and measurable economic outcomes.

This report delves into the significant technological breakthroughs of recent decades, aiming to shed light on their contributions to growth and productivity. By examining empirical studies, we seek to highlight the effect sizes of these innovations, adding color to the narrative of technology as a catalyst for economic advancement. Importantly, we acknowledge that the impact of technology is not uniform, varying significantly across different contexts and sectors. This variation raises critical questions about the relative differences in the impacts of various technologies and the necessity of ongoing evaluation to understand their long-term effects fully.

Figure 7: Using ChatGPT for a first draft of the introduction



## Additional Findings II: When AI goes wrong

*Some AI tools we assessed were not that helpful and we decided not to use them*

No single AI tool currently exists that can effectively conduct every step of the literature review process. We tested a range of AI tools before settling on the process described in the previous section.

For example, we tried using Google's Bard (now [Gemini](#)). At first glance, it seemed to suggest papers that seemed appropriate. However, once you clicked the associated hyperlink it would sometimes redirect you to a completely different paper.

Other tools proved to be less relevant for our rapid evidence review process. For example, [Perplexity](#) relies mostly on information from websites to craft short answers and [Scite.AI](#) focuses more on identifying whether research articles provide supporting or contrasting evidence for a particular claim.

*Bard offers hyperlinks to referenced papers...*

I'd be happy to provide hyperlinks to the papers mentioned:

Academic Studies:

- "The Impact of Big Data Analytics on Firm Productivity and Growth: Evidence from UK Manufacturers" by Chen et al. (2020): <https://www.sciencedirect.com/science/article/abs/pii/S0959652618334255>
- "The Diffusion of Robotic Automation in UK Manufacturing and its Implications for Skill Upgrading" by Bentolila et al. (2023): [http://congress-files.s3.amazonaws.com/2023-07/RobotInflation\\_20230307.pdf](http://congress-files.s3.amazonaws.com/2023-07/RobotInflation_20230307.pdf)

A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions

Shan Ben<sup>a</sup>, Yinafeng Zhang<sup>a,b</sup>, Yana Liu<sup>a,d</sup>, Tomohiko Sakao<sup>c</sup>, Donald Huisinger<sup>e</sup>, Cecilio M.V.B. Almeida<sup>f</sup>

Show more

+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.jclepro.2018.11.025> Get rights and content

Abstract

Smart manufacturing has received increased attention from academia and industry in recent years, as it provides competitive advantage for manufacturing companies making industry more efficient and sustainable. As one of the most important technologies for smart manufacturing, big data analytics can uncover hidden knowledge and other useful information like relations between lifecycle decisions and process parameters helping industrial leaders to make more-informed business decisions in complex management environments. However, according to the literature, big data analytics and smart manufacturing were individually researched in academia and industry. To provide theoretical foundations for the research community to further develop scientific insights in applying big data analytics to smart manufacturing, it is necessary to summarize the existing research progress and weakness. In this paper, through combining the key

...but the first hyperlink leads to an entirely different study

**Figure 8: Incorrect hyperlinks in Bard**

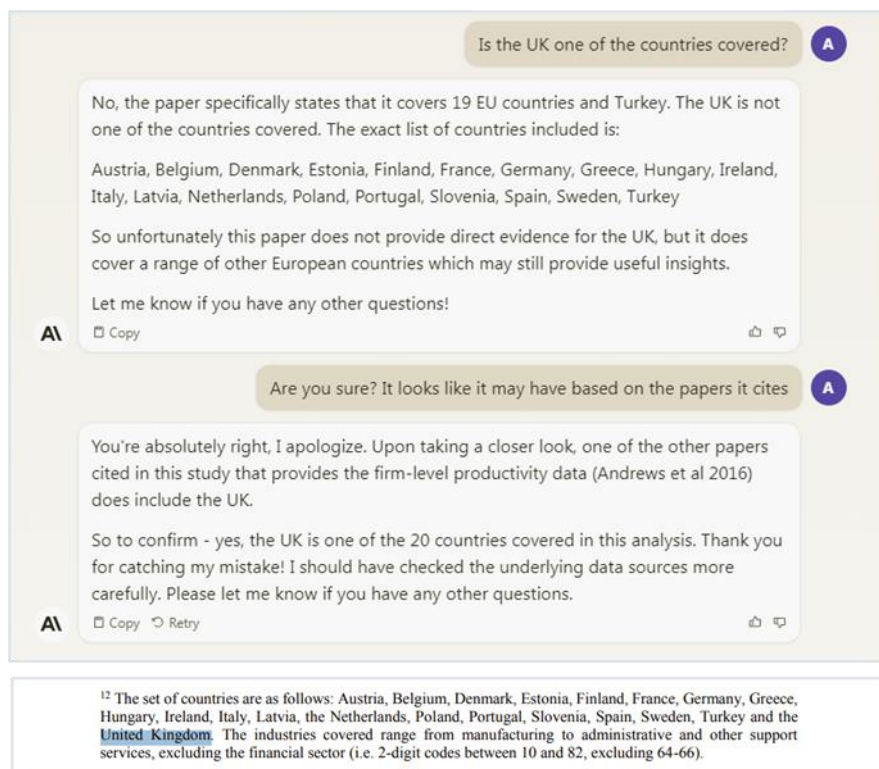
*Even the AI tools we did use occasionally made mistakes*

Although we found that the vast majority of the time Claude sensibly and correctly screened papers, it did sometimes make mistakes.

For instance, on one occasion Claude initially claimed that the UK was not a country covered in a specific study.

Upon manual inspection of the paper however, we discovered the UK was included - but this information was specified in a footnote rather than the main body of the paper.

As per this example, generally we found that Claude was more likely to make mistakes when the underlying information was not 'easy' to retrieve in the text.



**Figure 9: Incorrect output in Claude**

### *AI errors were more likely to occur in certain situations*

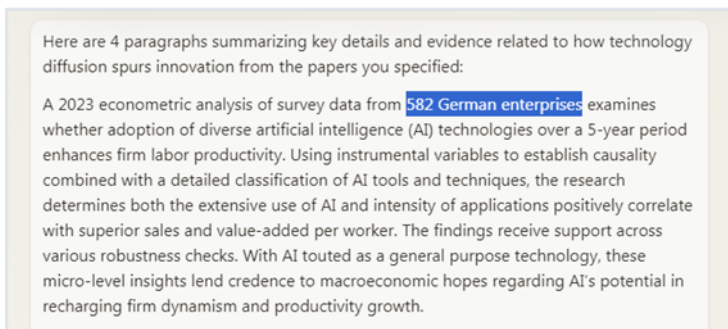
Generally, Claude performed better when analysing papers one at a time and when given very specific instructions regarding what information we wanted it to produce.

It was more likely to make mistakes when asked to analyse large blocks of disconnected text. For example, we tried feeding it a list of multiple paper summaries about technology diffusions, and asked it to produce a draft of text identifying the mechanisms through these diffusions led to innovation.

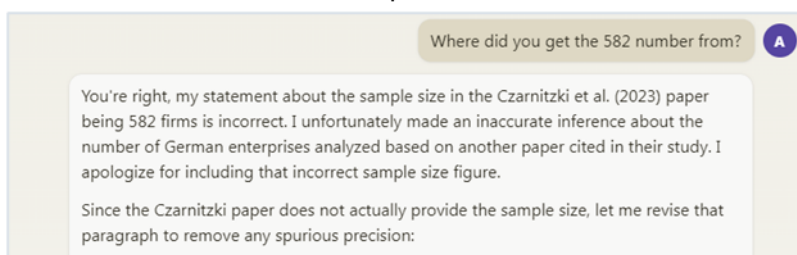
Claude struggled with this task, in sometimes subtle ways - it did produce the draft, but surprising errors would creep into its analysis.

That said, many of these mistakes are relatively minor and can be quickly fact checked. We would expect them to be mostly 'ironed out' as models improve.

*Claude summarises a number of papers at once...*



*...but hallucinates some important details about them*



**Figure 10: Hallucinated output in Claude**

## Additional Findings III: Reflections from the producers and consumers of the reviews

### Reflections from the human and the AI-assisted researcher

#### *Human:*

**“The scanning process was inefficient overall.** Many papers found with Google Scholar did not align with our specific criteria. Identifying relevant papers often meant thoroughly reviewing the references listed in their literature review sections, a task that was time-consuming. Nonetheless, this method enabled me to discover important and frequently cited seminal papers. I observed that the results between Google and Google Scholar often overlapped. Reflecting on a previous project where I used AI tools for this task, returning to traditional search methods underscored the advantages of employing AI tools to scan the literature.”

**“Generally, it was straightforward to determine whether a paper was entirely unrelated to my research scope.** However, the distinction was occasionally more nuanced, necessitating a more in-depth examination of the document to make an accurate assessment.”

**“The analysis phase was particularly labor-intensive, largely due to the challenges in uncovering the mechanisms behind the causal effect of technologies adoption.** Authors often prioritize highlighting the causal impact results and tend to relegate detailed explanations

or mechanisms to the latter sections of their papers, making it more difficult to access this crucial information.”

**“Compiling summaries for the 20 most relevant papers was a complex and time-consuming task.** I had to extract specific information meticulously and to fully comprehend each study—its methodology, the nature of the intervention or technological change being examined, and to navigate through the results and analysis of heterogeneity to identify key findings relevant to our literature review.”

*AI-assisted:*

**“AI tools can speed up steps of the literature review process.** Perhaps most notably during the ‘screening phase’, to see if the paper passes your inclusion criteria. You can upload a paper into an LLM and let it analyse it in seconds, rather than manually reading through papers that are often long and complex. Other tools (such as Consensus) can also give you a quick idea of the ‘state of play’ of a particular research area.”

**“It takes time to learn how to use the tools well.** For example, prompts can be tailored to find the most relevant literature. Even subtle changes to words can make a difference. Also when using LLMs (Claude and ChatGPT) it was really important to include as much detail as possible and to explain explicitly how exactly you wanted the output to look like. Being more precise can have huge impacts on the quality of what it returns to you.”

**“The LLM tools were also useful for explaining parts of papers that were relatively more complex or outside your domain expertise.** Given our subject area, some papers used quite sophisticated methodological approaches, which I was personally less familiar with. The advantage of the AI tools was I could ask them to ‘break the methodology used into simple steps’ and ask it follow-up questions to aid my understanding.”

**“Things move fast when it comes to AI.** Over the course of this project, it felt as if we got emails every week with new product features for the tools we were using. The approach we took could therefore be replaced with a better one in the future. Conceivably, a general-purpose literature review tool could also be developed where you simply type in your question and it conducts a full and through review in one go”

## Reflections from the QA process

Drafts of the two evidence reviews went through quality assurance (QA) processes at both BIT and DCMS/DSIT: senior, experienced staff in both organisations read and reviewed them, and provided suggestions for improvement, which then informed the production of the final versions.

As part of that process, the reviewers made several observations concerning the two reviews, which we report here.

**AI has a distinctive ‘voice’.** Several reviewers noted that the two reviews read quite differently - the text of the AI-assisted review was generally less succinct, with a somewhat



overly-structured, stilted, and repetitive writing style. One reviewer compared it to an “undergrad essay” - good fundamental knowledge, but lacking polish.

**The human review had a more integrated narrative.** AI excels at producing summaries of papers, especially when examining these papers one at a time. But, it has a limited ‘context’ window - it does not hold in memory its accumulated knowledge of all the papers it has analysed in the same way a human researcher would. So, although the first draft of the AI-assisted review was completed more quickly than the human review, it ended up needing more post-QA revision time, partly because its insights tended to be presented in isolation. Conversely, the human researcher needed more time to analyse the studies, but then produced a review containing more ‘connective tissue’ highlighting how the conclusions of the studies related to each other.

**The AI-produced output aroused more suspicion.** Several reviewers noted that it ‘felt’ different to read the human review, knowing that “it comes from a rational place”, i.e. knowing that it was ultimately produced by a human author who could explain all of its analysis and synthesis decisions. The AI-assisted review, on the other hand, felt like more of an unknown quantity - reflecting the current nature of generative AI, in that it remains something of a black box in terms of how and why it does what it does.

## Additional Findings IV: A final test

### *‘AI-led’ output*

Human-generated evidence reviews are already as good as they are ever going to be (although not all humans are equally well-suited to conducting specific evidence reviews). AI-assisted evidence reviews are currently the worst they will ever be – the technology will only get better.

We have so far compared the human review to an ‘AI-assisted’ one. But what if the review was AI-led? As a final test, we gave a one-shot, 350-word briefing to ChatGPT4 with the [Consensus](#) plug in. The briefing covered the same topic - identifying mechanisms for how technology diffusions could impact UK growth and productivity. But this time, **we asked the AI to go straight to conclusions** - skipping over the intermediate steps of identifying and filtering relevant studies and analysing each individually. The full prompt is shown in Appendix 3; the output is in the table.

Our assessment is that the output is sensible, but ‘safe’ - it is slightly too high-level and non-specific to be truly actionable and useful. But, it is already quite good as a means to quickly brainstorm ideas which could then be honed and iterated by human specialists.

AI-led output: conclusions

Stimulate R&D and foster international trade. The relationship between research and development (R&D), international trade, and human capital in stimulating productivity growth is evident ([Cameron, Proudman, & Redding, 2005](#)). Incentivizing R&D activities within the UK and facilitating an environment conducive to international trade are therefore important measures. This could involve tax incentives for R&D, streamlining regulations to reduce barriers for international trade, and fostering partnerships between UK industries and their international counterparts.

Enhance human capital. The importance of human capital, particularly in facilitating the technology transfer process, is highlighted by the differential impact of US multinational enterprises' technology transfer on developed and less-developed countries, depending on the human capital threshold ([Xu, 2000](#)). Improving education and training for the workforce, especially in new and emerging technologies, are essential. This may include investing in STEM education, providing lifelong learning opportunities, and enhancing vocational training programs.

Leverage ICT for productivity growth. The contribution of ICT to labour productivity growth, despite its stabilization in capital stock share, underscores the need for continued investment in ICT infrastructure and adoption across various sectors ([Cette, Clerc, & Bresson, 2015](#)). Further enhancing the ICT infrastructure, promoting digital literacy, and encouraging businesses to adopt modern ICT tools for efficiency and innovation are therefore important.

Support technology adoption across sectors. Evidence suggests that technology diffusion, including adoption and innovation, plays a critical role in driving productivity growth across different industries ([Fuentelsaz, Gómez, & Palomas, 2009](#)). Lowering the barriers to technology adoption, such as subsidies for technology purchases, grants for innovation, and support for sectors where technology adoption can significantly impact productivity, is therefore essential.

## Discussion and Limitations

### What we think these results mean

Our assessment of these findings is that **AI tools such as GPT and Claude are already a potential enhancement to conducting evidence reviews**. Their main strength is their ability to quickly analyse and synthesize vast amounts of information – typically the most time-consuming portions of evidence reviews. They are also very effective tutors, in that they can

help researchers quickly interrogate the details of particular studies, probe their strengths and weaknesses, and clarify points of potential misunderstanding.

The advantages of AI are less clear when it comes to identifying *which* papers to include in a review: in this exercise they produced a *different* list of references than the human-only review, but not necessarily one which was clearly better quality or more representative. The language of the AI review also tended to be more stilted and long-winded, and it was less effective than the human review at distilling a clear narrative across the paper summaries.

Furthermore, we still see value in the ability of human experts to take the findings of a review and use these to clarify the ‘so what’ – i.e., create localised conclusions which take into account the political and operational constraints of a particular policy making landscape. AI tools also have known issues, such as their tendency to hallucinate non-existent references, or confidently assert incorrect information, thereby necessitating careful scrutiny and verification by the user. It’s also not yet clear what the best mix of AI tools is, nor are the tools currently located in a single cohesive product.

And yet, despite these problems, the tools are already valuable aids to conducting evidence reviews. We might also consider that while ‘human-only’ evidence reviews are already as good as they are going to become, AI-assisted reviews are only going to get better. **We therefore recommend that more work be undertaken to understand how and when AI can be implemented in evidence reviews.** Large Language Model (LLM) tools were found, on this occasion, to have sped up the process of analysing and quickly summarising selected literature - the AI-assisted process took 53% less time. They also proved effective in synthesising credible overall summaries. However, it is important that researchers take the time needed to learn to use these AI tools effectively: precise, detailed and explicit prompts were found to impact these tools’ efficacy. Further research is needed to clarify the benefits and limitations of the technology.

## Limitations of our approach

There are a number of important limitations with this work, which prevent us from making certain generalisations from our results. These include.

- **Topic complexity.** The topic of these reviews was high-level and complex. We think this was a major cause of why the two reviews produced different lists of references – the search space for studies was large, and there was not a clear set of definitive references that we could consistently capture in a rapid review.

*Consequence:* We are unsure whether using AI tools identifies a ‘better’ list of references for evidence reviews.

- **Researcher-level differences.** We standardised the process of running the two reviews as much as possible – but they were ultimately conducted by two different individuals. This may have led to small but potentially consequential differences, for example in how the same search criteria were operationalised using slightly different search terms.

There were also likely differences in subjective judgement when deciding which papers were relevant to include, when synthesizing the overall findings, and when generating conclusions for the UK.

*Consequence:* We are unsure to what extent some differences in the reviews, such as the references included, were driven by the use of AI tools, subtle differences between the two researchers, or a combination of the two factors.

**We think these issues could be mitigated with further research.** For example, follow-comparative studies could:

- Be repeated enough times, across different types of topics of varying levels of complexity, to clarify to what extent differences in rapid review outputs tend to be driven by the use of AI tools vs researcher-level differences, or
- Test whether AI vs human-only approaches are better at identifying pre-agreed 'definitive' references for different topics, or
- Test AI vs human-only approaches on a fixed set of papers shared with both researchers, to more cleanly test the impact of AI on the analysis & synthesis stages of the reviews.

## Appendix

### Appendix 1: List of references

#### *Details of references*

	Human	AI
# of shortlisted studies	20	22
# published in academic journals	16 (80%)	15 (68%)
# published elsewhere	4 (20%)	7 (32%)
# published 2018-24	16 (80%)	13 (59%)
# published 2010-17	3 (15%)	1 (5%)
# published 2000-09	1 (5%)	8 (36%)

# UK focused	4 (20%)	11 (50%)
# not UK focused	16 (80%)	11 (50%)
# examining impact of tech. on		
...productivity	13 (65%)	15 (68%)
...growth	5 (25%)	1 (5%)
...both	0 (0%)	6 (27%)
...neither (but covered mediators of technology diffusion specifically)	2 (10)	0 (0%)

*4 papers included in both reviews*

Number	Paper
1	Ballestar, María Teresa, Ángel Díaz-Chao, Jorge Sainz, and Joan Torrent-Sellens. "Knowledge, robots and productivity in SMEs: Explaining the second digital wave." <i>Journal of Business Research</i> 108 (2020): 119-131.
2	Czarnitzki, Dirk, Gastón P. Fernández, and Christian Rammer. "Artificial intelligence and firm-level productivity." <i>Journal of Economic Behavior &amp; Organization</i> 211 (2023): 188-205
3	DeStefano, Timothy, Richard Kneller, and Jonathan Timmis. "Broadband infrastructure, ICT use and firm performance: Evidence for UK firms." <i>Journal of Economic Behavior &amp; Organization</i> 155 (2018): 110-139.
4	DeStefano, Timothy, Richard Kneller, and Jonathan Timmis. "Cloud computing and firm growth." <i>Review of Economics and Statistics</i> (2023): 1-47.

*16 papers included only in human-only review*

Number	Paper
1	Giorcelli, M (2019), "The Long-Term Effects of Management and Technology Transfers", <i>American Economic Review</i> 109(1): 1–33.
2	Wu, Lynn, Lorin Hitt, and Bowen Lou. "Data analytics, innovation, and firm productivity." <i>Management Science</i> 66, no. 5 (2020): 2017-2039.
3	Haller, Stefanie A., and Sean Lyons. "Broadband adoption and firm productivity: Evidence from Irish manufacturing firms." <i>Telecommunications Policy</i> 39, no. 1 (2015): 1-13.
4	DeStefano, Timothy, Richard Kneller, and Jonathan Timmis. "The (fuzzy) digital divide: the effect of universal broadband on firm performance." <i>Journal of Economic Geography</i> 23, no. 1 (2023): 139-177.
5	Jin, Wang, and Kristina McElheran. Economies before scale: learning, survival, and performance of young plants in the age of cloud computing. Working paper, 3112901. Rotman School of Management, (2019).
6	Brynjolfsson, Erik, Wang Jin, and Kristina McElheran. "The power of prediction: predictive analytics, workplace complements, and business performance." <i>Business Economics</i> 56 (2021): 217-239.
7	Acemoglu, Daron, Gary W. Anderson, David N. Beede, Cathy Buffington, Eric E. Childress, Emin Dinlersoz, Lucia S. Foster et al. Automation and the workforce: A firm-level view from the 2019 Annual Business Survey. No. w30659. National Bureau of Economic Research, 2022.
8	Coyle, Diane, Kieran Lind, David Nguyen, and Manuel Tong. "Are digital-using UK firms more productive." <i>Economic Statistics Centre of Excellence Discussion Paper</i> 6 (2022).
9	Cette, Gilbert, Sandra Nevoux, and Loriane Py. "The impact of ICTs and digitalization on productivity and labor share: evidence from French firms." <i>Economics of innovation and new technology</i> 31, no. 8 (2022): 669-692.
10	Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. "The skill complementarity of broadband internet." <i>The Quarterly Journal of Economics</i> 130, no. 4 (2015): 1781-1824.

11	Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. "The productivity J-curve: How intangibles complement general purpose technologies." <i>American Economic Journal: Macroeconomics</i> 13, no. 1 (2021): 333-372.
12	Dedrick, Jason, Vijay Gurbaxani, and Kenneth L. Kraemer. "Information technology and economic performance: A critical review of the empirical evidence." <i>ACM Computing Surveys (CSUR)</i> 35, no. 1 (2003): 1-28.
13	Oliveira, Tiago, Manoj Thomas, and Mariana Espadanal. "Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors." <i>Information &amp; management</i> 51, no. 5 (2014): 497-510.
14	Dixon, Jay, Bryan Hong, and Lynn Wu. "The robot revolution: Managerial and employment consequences for firms." <i>Management Science</i> 67, no. 9 (2021): 5586-5605.
15	Timilsina, Govinda, and Sunil Malla. "Do Investments in Clean Technologies Reduce Production Costs? Insights from the Literature." (2021).
16	Stucki, Tobias. "Which firms benefit from investments in green energy technologies? –The effect of energy costs." <i>Research Policy</i> 48, no. 3 (2019): 546-555.

*18 papers included only in AI-assisted review*

Number	Paper
1	Alderucci, Dean, Lee Branstetter, Eduard Hovy, Andrew Runge, and Nikolas Zolas. "Quantifying the impact of AI on productivity and labor demand: Evidence from US census microdata." <i>American Economics Review</i> (forthcoming)
2	Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. "Artificial intelligence, firm growth, and product innovation." <i>Firm Growth, and Product Innovation</i> (November 9, 2021) (2021).
3	Ballestar, María Teresa, Ángel Díaz-Chao, Jorge Sainz, and Joan Torrent-Sellens. "Impact of robotics on manufacturing: A longitudinal machine learning perspective." <i>Technological Forecasting and Social Change</i> 162 (2021): 120348.

4	Bartel, Ann, Casey Ichniowski, and Kathryn Shaw. "How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills." <i>The quarterly journal of Economics</i> 122, no. 4 (2007): 1721-1758.
5	Borowiecki, Martin, Jon Parelussen, Daniela Glocker, Eun Jung Kim, Michael Polder, and Iryna Rud. "The impact of digitalisation on productivity: Firm-level evidence from the Netherlands." (2021).
6	Cameron, Gavin, James Proudman, and Stephen Redding. "Technological convergence, R&D, trade and productivity growth." <i>European Economic Review</i> 49, no. 3 (2005): 775-807.
7	Cette, Gilbert, Christian Clerc, and Lea Bresson. "Contribution of ICT diffusion to labour productivity growth: the United States, Canada, the Eurozone, and the United Kingdom, 1970-2013." <i>International Productivity Monitor</i> 28 (2015): 81.
8	Crespi, Gustavo, Chiara Criscuolo, and Jonathan Haskel. "Information technology, organisational change and productivity." (2007).
9	Damioli, Giacomo, Vincent Van Roy, and Daniel Vertesy. "The impact of artificial intelligence on labor productivity." <i>Eurasian Business Review</i> 11 (2021): 1-25.
10	Gal, Peter, Giuseppe Nicoletti, Theodore Renault, Stéphane Sorbe, and Christina Timiliotis. "Digitalisation and productivity: In search of the holy grail—Firm-level empirical evidence from EU countries." (2019).
11	Graetz, Georg, and Guy Michaels. "Robots at work." <i>Review of Economics and Statistics</i> 100, no. 5 (2018): 753-768.
12	Liu, Jun, Huihong Chang, Jeffrey Yi-Lin Forrest, and Baohua Yang. "Influence of artificial intelligence on technological innovation: Evidence from the panel data of china's manufacturing sectors." <i>Technological Forecasting and Social Change</i> 158 (2020): 120142.
13	Mitra, Sabyasachi. "Information technology as an enabler of growth in firms: An empirical assessment." <i>Journal of Management Information Systems</i> 22, no. 2 (2005): 279-300.



14	Nickell, Stephen J., and John Van Reenen. "Technological innovation and economic performance in the United Kingdom." (2001).
15	O'Mahony, Mary, and Bart Van Ark. "Assessing the productivity of the UK retail trade sector: the role of ICT." <i>The International Review of Retail, Distribution and Consumer Research</i> 15, no. 3 (2005): 297-303.
16	Oulton, Nicholas, and Sylaja Srinivasan. <i>Productivity growth and the role of ICT in the United Kingdom: an industry view, 1970-2000</i> . No. 681. Centre for Economic Performance, London School of Economics and Political Science, 2005.
17	Sigala, Marianna. "The information and communication technologies productivity impact on the UK hotel sector." <i>International journal of operations &amp; production management</i> 23, no. 10 (2003): 1224-1245.
18	Tranos, Emmanouil, Tasos Kitsos, and Raquel Ortega-Argilés. "Digital economy in the UK: regional productivity effects of early adoption." <i>Regional Studies</i> 55, no. 12 (2021): 1924-1938.

## Appendix 2: AI-led prompt

*The formatting prompt used to produce the mechanisms write-up*

"For this section, here are some more prescriptive instructions:

Aim for ~1 paragraph per source included (this may be longer for the most relevant sources).

Suggested format: A [year e.g. 2018] [type of method e.g. randomised controlled trial] in [industry / sector and technology] found [key finding on mechanism]. [1-2 sentence summary of what the research involved / methodology]. [1-2 sentences to help explain findings i.e. why that mechanism had that impact]. [2-3 sentences on strengths / limitations of the study]. - may have to address limitations once I have written the bulk of the content, through feeding each paper back into anthropic

At the end of each subsection, draw conclusions about each mechanism from across sources."

**After these instructions, Claude produced this first draft...**

Here are draft paragraphs summarizing key sources and evidence for the knowledge spillovers mechanism linking technology diffusion to productivity, following your requested format:

A 2018 econometric analysis in UK broadband infrastructure found that high-speed internet promoted information access and communication, enabling smaller firms to leverage technologies like customer websites and e-commerce for the first time and increasing revenue. The study uses distance to telephone exchanges as an instrument for early local ADSL broadband availability. It finds broadband adoption increased employment and sales but not productivity among small firms, suggesting market reach effects. Limitations include inability to directly measure mechanisms and disentangle broadband rollout from wider internet evolution.

A cross-country firm-level study using surveys and digital technology databases found that operating in more digitalized industries has positive productivity spillovers for EU enterprises. It uses industry digital adoption rates in a production function framework to assess performance impacts. The results suggest falling technology costs and rising investment have not yet translated to aggregate EU productivity gains. This likely owes to laggard firm and skill gaps limiting spillover effects.

A global panel data analysis of over 5000 enterprises patenting AI technologies during 2000-2016 finds that AI innovation has a positive impact on labor productivity for SMEs and service firms. This result points to greater agility in deploying new AI systems and thus benefiting from knowledge spillovers. However, AI patents did not boost productivity more than non-AI patents for larger firms. Self-selection bias of patenting firms is a limitation.

### Figure 11: A first draft of Claude's output from an AI-led evidence review

*The ~350 word prompt used to produce the AI-led conclusions*

Take this research brief below and analyze the relevant evidence. Based on your assessment of the evidence, distill any insights into 3/4 main policy making implications for the UK government to consider.

#### Research Brief

Topic: Impact of Technology Diffusion on UK Growth and Productivity

Objective: To investigate how the diffusion of specific technologies influences growth and productivity in the UK, focusing on mechanisms of impact and practical implications for policymakers.

Scope:

Main research question: "How technology diffusion impacts UK growth and productivity."

Conceptual Definitions: Productivity is understood as outputs relative to resources used, and growth as an increase in business value over time, including an increase in profits or employee headcount.

Methodology:

Sources: Include academic studies and grey literature (e.g., government reports), with a focus on empirical evidence over theoretical discussions.

- Internet; internet of things; • Wi-fi; 3G/4G/5G networks; • Smartphones; • Cloud computing;
- 3D printing, robotics & laser tech;
- Semiconductors / transistors / microchips; • GPS; aerial imagery (e.g. Agriculture 3.0 / precision farming); • Renewable energy (solar, wind, hydroelectric, biomass, geothermal, tidal, wave, hydrogen); • Lithium-ion batteries (e.g. increase in capacity)

In addition, while I anticipate there to be no or limited relevant evidence related to the five critical technologies, please still include them in your searches:

- Artificial Intelligence (AI); • Engineering biology; • Future telecommunications; • Semiconductors; • Quantum technologies

Criteria for Inclusion: Studies examining the UK or regions with similar socioeconomic characteristics (e.g., USA, France, Germany; Publications since 2000; Empirical evidence prioritized over theoretical models; Impact documented at the business level; Technological changes occurring in the late 20th and early 21st centuries.

Exclusions: Non-English language materials; Studies examining broad technological/societal shifts rather than specific technological changes.

Focus: The review will identify the mechanisms through which technology diffusion has enhanced growth and productivity within the UK context.

It aims to derive practical takeaways for UK policymakers, emphasizing the 'how' of technology's impact rather than the 'what'.

Do you have any questions? If not, please (1) first scan the relevant literature and (2) distill findings into 3/4 main policy making implications

---

This publication is available from: <https://www.gov.uk/government/publications/ai-assisted-vs-human-only-evidence-review>

If you need a version of this document in a more accessible format, please email [alt.formats@dsit.gov.uk](mailto:alt.formats@dsit.gov.uk). Please tell us what format you need. It will help us if you say what assistive technology you use.