Using Generative Artificial Intelligence (GenAI) across different Research Phases – Cases, Potential and Risks

Report to the Danish Council for Research and Innovation Policy (DFiR) by the Danish Centre for Studies in Research and Research Policy (CFA), Aarhus University, September 2024

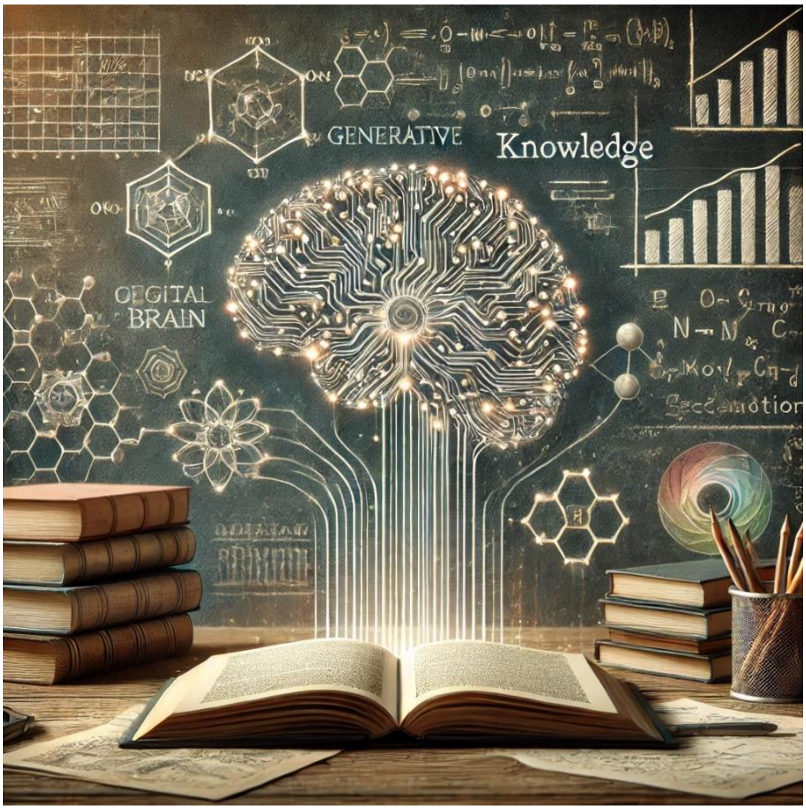


*Image generated by ChatGPT (Generative AI and the Research Ecosystem)

Authors:
Mads P. Sørensen, Professor, Danish Centre for Studies in Research and Research Policy, Aarhus University
Serge P. J. M. Horbach, Assistant Professor, Institute for Science in Society, Radboud University
Oksana Dorofeeva, PhD student, Danish Centre for Studies in Research and Research Policy, Aarhus University
Mads Schäfer Bak, PhD student, Danish Centre for Studies in Research and Research Policy, Aarhus University

DANISH CENTRE FOR STUDIES IN RESEARCH AND RESEARCH POLICY
DEPARTMENT OF POLITICAL SCIENCE
AARHUS UNIVERSITY

Executive summary

Generative Artificial Intelligence (GenAI) has seen rapid development in recent years, including in areas affecting research. GenAI models and tools can now perform sophisticated research tasks such as analysing large datasets, summarizing extensive text information, and generating synthetic data. This makes GenAI interesting for the research sector and it forces the research community to critically examine its current ways of doing research and thinking about the research process.

This report examines the use of GenAI in the research process, across different Research phases. It looks at various use cases and the potential and risks in using GenAI in the research process. The report begins by defining GenAI and looking at the evolutionary progress of various GenAI models and tools. Hereafter, GenAI's use in different phases of the research process is examined using a five-phase analytical model, consisting of the following phases: Idea generation and research funding, Research design, Data collection, Data analysis, and Scientific publishing, reporting, and dissemination.

The report is based on an extensive literature review, complemented by expert interviews. It shows that most research fields can use GenAI in the research process for tasks such as idea generation, literature reviews, translation, language editing, and writing. GenAI tools can, for example, assist researchers in drafting grant applications or preparing manuscripts for publication. However, when it comes to data collection and analysis, there is considerable variation in the relevance and utility of GenAI tools across disciplines and research fields. Differences in the research areas, epistemic cultures, and methodological approaches influence how researchers in different disciplines can benefit from GenAI. Fields that rely heavily on qualitative analysis or human interpretation may, for example, find GenAI less useful for data analysis than other fields.

The report concludes that there is no one-size-fits-all solution for integrating GenAI into the research process. Strategies for adopting GenAI need to be tailored to the specific needs and contexts of individual research fields. To figure out what works best within a particular field, a collaborative approach is recommended, where researchers and epistemic communities along with relevant stakeholders such as funders and publishers work together to develop policies and guidelines that maximize the benefits of GenAI while minimizing risks. These collaborative efforts could involve roundtable discussions and stakeholder consultations to ensure that strategies are as relevant and effective as possible.

It is also crucial to ensure that the increased speed of knowledge production enabled by GenAI does not compromise the quality of the research produced. Although GenAI can help accelerate various research processes, care must be taken to maintain rigour and thoroughness in research methods and outputs. The acceleration of research may also put pressure on other processes, such as peer review or academic reward systems. These potential impacts need to be carefully considered to avoid unintended consequences in the academic ecosystem. Ultimately, finding a balance between the speed and efficiency offered by GenAI and the need for high-quality, reliable research outcomes is essential for its successful integration into the research process. Moreover, the benefits of utilising GenAI within research must be balanced with the risks, including the pressure this technology puts on planetary resources.

1. Introduction

Generative Artificial Intelligence (GenAI)'s capacity to generate text and images – as well as music, videos and other types of output – has developed dramatically in recent years. This development has especially been led by private companies such as Microsoft, Google, Bing and OpenAI with its flagship model, ChatGPT, released to the public in November 2022. The development of Large Language Models (LLMs), like ChatGPT, and other GenAI models such as GANs (Generative Adversarial Networks) has led to the expansion of GenAI-based tools that are now able to perform challenging tasks such as analysing huge datasets, summarizing vast amounts of information, generating comprehensive reports, creating images, and generating other valuable outputs.

This development that has taken place within just a few years also raises important questions about GenAI's implications for the research sector. The advanced state that GenAI has now reached makes it highly interesting for the research sector and it forces the research community to critically examine its current ways of doing research and thinking about the research process (see e.g. The Royal Society, 2024; Cornell University Task Force, 2023).

The examination and rethinking of research practices have already begun in research institutions throughout the world. Across the different phases of the research process, researchers from all main areas of research have started experimenting with using GenAI to support their research work.

As we shall see in this report, GenAI is already today used for tasks ranging from helping researchers generate hypotheses and identifying the most relevant literature to developing research designs, and collecting, generating and analysing data. It is also used for research applications and for other writing and dissemination tasks. However, when examining the potential of GenAI, it is important to also recognize potential risks related to the widespread use of GenAI in research. These risks include concerns related to research ethics, research integrity, and the environmental impact of GenAI.

The present report discusses various examples of GenAI use in different research contexts across the entire research process. We begin in Chapter 2 by defining GenAI and tracing its evolution from early AI concepts to the advanced GenAI models we see today, such as LLMs like those developed by OpenAI. The methodology chapter (Chapter 3) outlines the approach taken to gather data and insights for the report, which includes a comprehensive literature review on GenAI use cases supplemented with individual interviews with researchers based in Denmark about their experiences with GenAI and how it is or could be used within their research fields.

Following this, we introduce an analytical model for the research process in Chapter 4 consisting of five phases: a) Idea generation and research funding, b) Research design, c) Data collection, d) Data analysis, and e) Scientific publishing, reporting, and dissemination. After introducing the model, we examine each of the five phases, providing an overview of the use of GenAI in each phase and discussing related potential and risks. We use both examples from the literature review and the interviews to show how GenAI is today already integrated into the different phases of the research process and to discuss opportunities and limitations for future use.

Next, in Chapter 5, we sum up and discuss the potential benefits and inherent risks of using GenAI in research. In Chapter 6, we present the key results of CFA's recent survey of Danish researchers, which provides insights into current attitudes and practices regarding GenAI usage in the Danish research community. Finally, Chapter 7 summarizes and discusses the main findings of the report, including its limitations.

The report is the result of a commissioned study for The Danish Council for Research and Innovation Policy (DFiR) carried out by the Danish Centre for Studies in Research and Research Policy (CFA). The study took place over the summer and early autumn of 2024.

2. Definitions and historical development of Generative Artificial Intelligence (GenAI)

2.1. Defining generative artificial intelligence

Generative artificial intelligence (GenAI) is a term for a type of artificial intelligence created within the currently dominant AI approach, machine learning. The 'generative' indicates that this class of machine learning models generates or creates something as its output. For instance, GenAI-based tools like Midjourney and DALL-E generate new images based on textual prompts. However, this terminology alone cannot distinguish GenAI models from non-generative ones, as all models produce something new, like rules or predictions (García-Peñalvo and Vázquez-Ingelmo, 2023, p. 7). A simple regression model generates a new linear relationship between two variables. Focusing on the popular tools and use cases, both the wider public perceptions and existing definitions of GenAI emphasize that GenAI models create human-like content (García-Peñalvo and Vázquez-Ingelmo, 2023), like text (LLMs), images, music, etc.

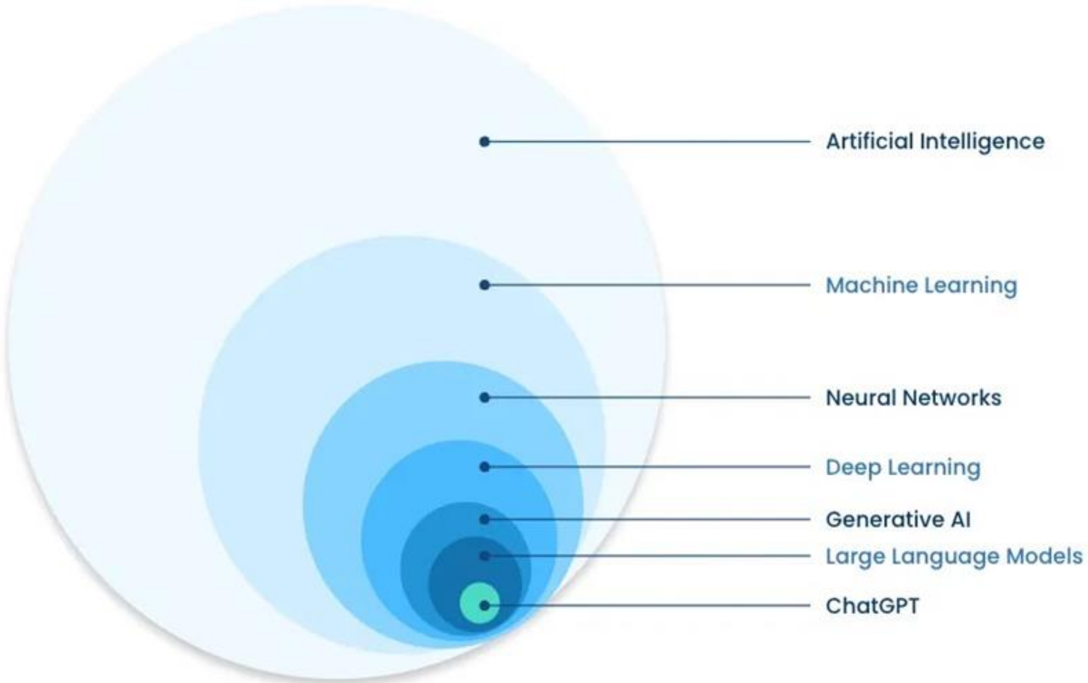
However, defining GenAI as a class of models that produce content is limiting, especially when considering the context of science and research. More technically accurate would be to say that GenAI models produce new data samples based on the patterns of the data they were trained on (Bengesi et al., 2024, p. 69813). These new data samples are conditioned by and contingent on the data they were trained on, so models based on content data (text, music, etc.) would produce new content. Depending on what data they are trained on, GenAI models can produce various outputs we do not associate with content, such as new tabular data, protein formulae, etc. Models with these generative capabilities differ in their inner workings from other machine learning models.

2.2. Situating generative AI: history and position within the broader AI field

Generative models are a subset of machine learning, a subfield of artificial intelligence focusing on creating and deploying models that 'learn' to perform tasks in a way that we associate with human intelligence: by induction and learning from examples rather than following strict rules (see Figure 2.1. for a visualisation of AI subfields). Particularly, neural networks are modelled after the way organic brains process information. Artificial neurons (analogy of brain neurons) are aggregated into layers, from the input to the output layer, which comprise the architecture of the neural network. Neurons in each layer receive signals, transform them, and send them to the next layer. The large number of layers distinguishes deep learning from traditional machine learning approaches. This approach allows models to capture patterns in the data without being given strict instructions. For instance, a neural network, given examples of two classes of images (e.g., apples and oranges), learns to discriminate between these two classes and predict whether a new image depicts an apple or an orange.

Such discriminative models are mature technology: most commonly used machine learning techniques, including many deep learning models, come from the late 20th century (Bengesi et al., 2024). In the 2010s, significant increases in data and computing resources facilitated a breakthrough in deep learning: the discriminative models got increasingly complex, demonstrating impressive results in tasks from many research areas. A famous example is the skin cancer diagnosis method reported to perform at the level of medical doctors (Esteva et al., 2017). These machine learning models have had a tremendous effect on research in many areas of science. Supervised machine learning is widely used in physics for identification of known particles, remote sensing research in the natural sciences, etc. (Karagiorgi et al., 2022; Ali et al., 2015).

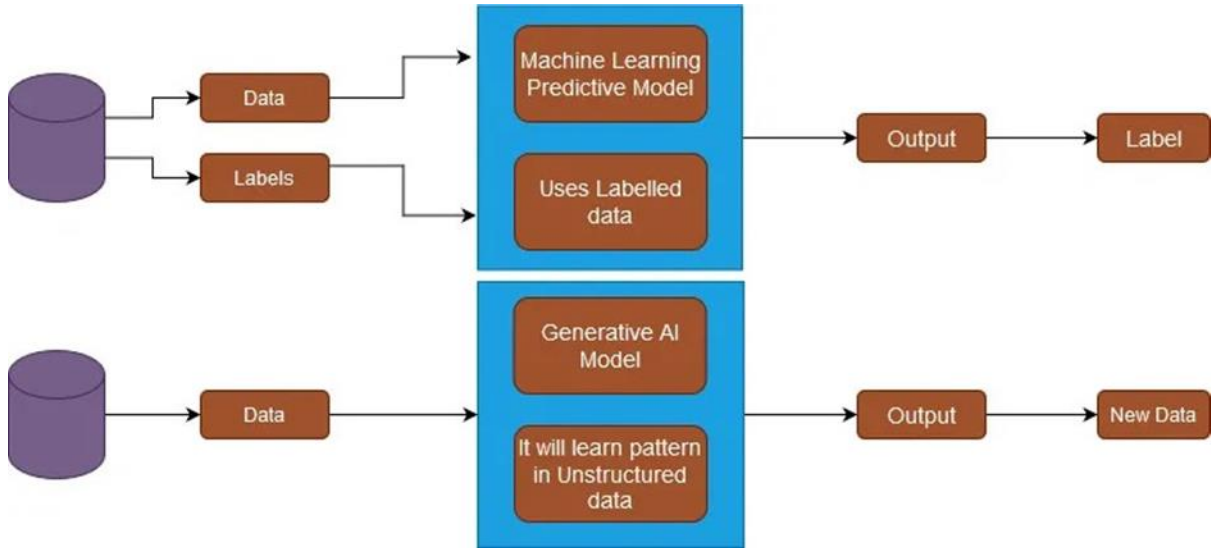
Figure 2.1. Generative AI within the Artificial Intelligence field (Stefano, 2023)



This deep learning breakthrough of the 2010s also brought about the invention and fast development of generative deep learning models. Whereas discriminative models learn the decision boundary separating the classes (e.g., distinguishing between apples and oranges) to predict in which class a new sample would fall, generative models learn the underlying probability distribution of the whole (often large) dataset. They then recreate this probability distribution, based on which they generate a new entity.

So, generative models would learn all the features of apples and oranges to generate new images of the fruit. Another way to distinguish them would be through the idea of supervision: prediction models would be 'supervised' by the provided labels (e.g., this is an orange, this is an apple) while generative models draw on unsupervised (or semi-supervised) learning, which means a model learns not from the labels but from the underlying distribution, as illustrated in Figure 2.2.

Figure 2.2. Discriminative/predictive models vs. generative models (Sharma, 2024)



2.3. Generative models have been around for a long time but exploded in mid-late 2010s

Generative computational methods existed before this moment: they date back to the mid-20th century and were steadily progressing within areas such as natural language processing (NLP) and computer vision (Cao et al., 2023). However, only the advent of deep learning-based generative models in the 2010s has brought the performance of generative methods to an exciting level. In 2013, Diederik Kingma and Max Welling introduced the concept of Variational Autoencoders (VAEs), and in 2014, Ian Goodfellow and his colleagues invented a new type of neural network: generative adversarial networks (GANs). Thus came two of the (now) main foundations of GenAI (Bengesi et al., 2024). From then on, the term 'generative artificial intelligence' started to spread, and generative models began fast development (Bengesi et al., 2024; García-Peñalvo and Vázquez-Ingelmo, 2023). The next revolution in the development of GenAI came from the 2017 paper 'Attention is all you need' by Ashish Vaswani and colleagues from Google (Vaswani et al., 2017). They proposed a concept of attention and an architecture for neural networks based around the attention mechanism. Attention is a mechanism that allows the model to focus on relevant parts of the input data; the transformer architecture employs this mechanism multiple times to create a dynamic focus on multiple parts in the input at the same time instead of processing data in a linear fashion (Khoshkebari, 2023). Khoshkebari (2023) compares this functionality to reading a mystery novel in a way that simultaneously captures all the relevant clues to identify the culprit and pieces the solution together. This technology allowed models to analyse extra-large datasets and to train models that combine data from different modalities (e.g., text and images) (Bengesi et al., 2024; Cao et al., 2023). The transformer architecture, for instance, forms the basis for famous GPT models, as indicated by their name (GPT stands for Generative Pre-trained Transformer) and the AlphaFold2 method that facilitated a breakthrough in the scientific field of protein prediction (Cao et al., 2023; Moussad et al., 2023). The three aforementioned model types, VAEs, GANs and transformer models, together with diffusion models, and variants and combinations of the four, constitute the current GenAI landscape (Bengesi et al., 2024; García-Peñalvo and Vázquez-Ingelmo, 2023).

2.4. 'Generative AI' does not always mean the same thing

A literature study done by García-Peñalvo and Vázquez-Ingelmo (2023) found that the term 'generative AI' is mostly featured in editorials or papers that discuss the impact of GenAI tools on various domains. On the other hand, the AI research community discussed the specific generative models (GANs, transformers, etc.) without marking this work as 'generative AI'. Thus, they connect this term to the discourse that started with the release of generative model-based commercial tools like ChatGPT. Indeed, the public release of the OpenAI product has led to an explosion of publications related to GenAI.

While this reported divergence in terminology might have decreased due to the continued attention to GenAI, it remains important to consider for a more comprehensive view of the GenAI landscape as well as for facilitating a dialogue on GenAI between technical and non-technical communities. Moreover, the technology hype around 'generative AI' makes a buzzword out of the term: using it might add value on its own (e.g., attracting attention, excitement, prestige, investment, etc.), regardless of the benefits or downsides of actually using GenAI. That incentivizes the use of the term, both in the media and in academic works, and thus can widen and/or blur its definition.

3. Literature review and interview methodology

This report is based on two complementary streams of data: an overview of the academic literature on the use of GenAI and interviews with selected researchers about their use and interpretation of GenAI and its future potential and risks for academic research.

3.1. Literature review

The systematic literature review was conducted using the Web of Science database, by means of the Web of Science Core Collection. The search string was developed in an iterative process aiming to capture a wide variety of use cases of GenAI across various stages of the research process and diverse research disciplines and knowledge production ways. The final search string was designed in such a way that it mainly captures articles that describe either new applications of GenAI models, GenAI tools for academic research, or methods of incorporating GenAI in research efforts. That is to say, the search string is tailored to pioneering articles that either suggest or evaluate such new applications, rather than studies that have implicitly integrated these applications into parts of their research methods. The full search string can be found in appendix 9.1. The search was limited to publications in the last five years (starting 2019 until the search date) and to English language documents.

In developing the search strategy and search string, both academic and grey literature were considered. Because of the large, and rapidly growing, academic literature on the topic covering all relevant aspects of our study and the better accessibility of this literature compared to grey literature, the final search was conducted on academic research only. Nevertheless, the grey literature consulted when developing the search strategy has also been incorporated in the remainder of the analysis and is referenced accordingly in this report.

The search was initially conducted on July 4, 2024, and subsequently repeated to capture the latest research material on August 4, and September 4, 2024. This resulted in 601, 688 and 753 research articles respectively. This indicates the fast-moving nature of the research front related to GenAI. It took 5 years to reach 600 articles, but in the last two months this number has increased by 25%, with 2.5 new articles appearing on this topic every day just within the rather narrow scope of our search string. It also indicates that any overview of GenAI for research purposes necessarily only provides a snapshot of the current state of play, which is likely to quickly develop, potentially also into unexpected territory.

Scope

In our search and in the remainder of this document, we focus on researchers aiming to use GenAI models and applications to contribute to research in productive and positive ways. We explicitly do not include applications of GenAI for malicious purposes, such as the creation of papermills, data fabrication or other forms of research misconduct, even though we acknowledge that such malicious use is possible and potentially constitutes a major threat to the research community.

In addition, our research focuses on the use of GenAI in academic research contexts. We are aware that many interesting developments regarding (generative) artificial intelligence are fuelled by non-academic actors, including industry and governmental actors, but these are out of the scope of this report. Similarly, the report focusses on research-related tasks within academia, thereby not explicitly covering the use of GenAI for teaching purposes, although some applications described in this report also have educational aspects to them.

Data extraction

Based on the final set of 753 documents, we extracted the use cases of GenAI for research purposes described in them, keeping track of the research fields for which this use case is relevant and the stage of the research process to which it applies. Note that some of these articles merely propose a use case of GenAI, whereas others contain proofs-of-concepts, showing how these have been used in practice. We also extracted potential benefits and risk associated with the use case, whenever these were described or discussed in the document. When building our repository of use cases, we took a saturation approach. That is to say that we included use cases that were genuinely different from the ones that were already in our repository, omitting use cases that were (nearly) identical, e.g. descriptions of the same use case in a different research field. Consequently, some of the use cases presented below may also apply to research contexts different from the ones we describe.

3.2. Expert interviews

In addition to the literature review, we conducted 9 interviews with researchers in Denmark. Complementing the literature review with the perspective of researchers provides a more nuanced picture of GenAI in research. First, it gives examples of the current real-life cases of GenAI use in research carried out in Denmark. Secondly, it gives insight into the experiential side of working with GenAI: what works well in practice versus what does not. It also allows insight into researchers' motivations for (non-)use of GenAI in various scenarios.

The selection of experts contributing to this report balances the following considerations. First, we wanted to include some researchers actively working with GenAI to provide examples of successful use cases. For that, we looked for researchers authoring papers and talks on GenAI/generative modelling or researchers featuring in media articles about research using GenAI. The latter yielded mixed results as the media sometimes reported the use of GenAI but further investigation (looking at the academic outputs or talking to the researcher in question) revealed that it was not GenAI, but a more traditional (classification) deep learning used. These cases illustrate the murkiness of the terminology around (generative) AI, particularly in popular discourse, and the role of hype in that (cf. Chapter 2). We also aimed to include experts whose research focuses on GenAI, including in terms of the biases, trustworthiness, etc. of these technologies. Moreover, we strived to represent a diverse range of research areas among our expert contributors. Gender equality was also a consideration in our sampling strategy. In addition to these strategies, we benefited from recommendations from multiple researchers, including the experts contributing to this report and the researchers who were approached but for one reason or another had to decline our invitation.

In total, 24 experts were contacted via email (20 Denmark-based researchers and 4 experts from abroad); this resulted in 9 participants (all based at Danish institutions). They were selected as experts within their research fields, who have experience with AI. This could be as advanced users and/or in some cases as experts who study AI, incl. GenAI. The expert interviews were used to collect actual use cases and to gain insight into the potential and risks of applying generative models in the research process. The gender balance was skewed both in the search phase (based on how many relevant experts we could identify) and in the final corpus. 8 female researchers were contacted, resulting in the recruitment of only one. Brief biographies of the participating experts are included in appendix 9.2. One expert asked to not be included in the report.

The interviews focused on learning about the experts' personal experiences with GenAI across research phases and their perspectives on the risks and benefits of such practices. Moreover, we also invited them to reflect on the use of GenAI in their field, the incentives and barriers for (non-)use of GenAI in their field, as well as future developments.

In addition, the interviews touched upon the topics of training and policies on GenAI in research. For more details on the themes covered in the interviews, we refer to the interview guide in appendix 9.3.

The interviewees participate in this report in their capacity as experts within their fields; their contributions are acknowledged in this report, and the quotes from interviews are attributed to each expert. Before the interview, we informed the experts about participation details and our processing of their personal data. Written informed consent was obtained from all interviewees. The interviews were conducted online via Microsoft Teams and videorecorded; for analysis, we worked with the audios and auto-generated transcripts.

4. The use of GenAI across disciplines and research phases

In this chapter we first present the five phase-model of the research process that we will use to present and discuss different cases of GenAI use in research (Chapter 4.1). Hereafter, we go through each of the five phases, one by one, and present examples of how GenAI is used to support the work within each of these phases of the research process (this chapter, sections 4.2-4.6). The examples are collected from the literature review and from the expert interviews. They cover different disciplines and areas of research and reflect a multitude of knowledge production ways. Each section is supplemented with a table that provides further examples of GenAI use within a particular phase. The tables describe the use cases as well as refer to their sources and the field/discipline context within which they are reported.

4.1. The Research Process and the Five-phase model

There are various approaches to producing knowledge within the research community, leading to significant differences in the research processes that underpin the knowledge produced. The research process does not just vary between disciplines, but also within disciplines and main areas of research. In the social sciences, for example, a prominent divide is between qualitative and quantitative research. Although many social science studies today combine qualitative and quantitative methods, work with a multi-methods or mixed-methods approach, the distinction between qualitative and quantitative research remains present in the social sciences, and social scientists are typically specialized in one or the other.

In other main areas of research, we similarly find diversity in the way knowledge is produced. In the medical, biomedical and health sciences there is a division between basic research and clinical/relational research, in the natural sciences between laboratory research and theoretical research, and in the technical sciences between experimental/laboratory research and theoretical research. In the humanities, one can make similar distinctions, e.g., between research done on existing data/data produced by others (i.e. texts, paintings, movies, ideas, thought systems, religions, media material etc.) and research on new data collected by the researchers themselves (i.e. interview material, ethnographic data, survey data, archaeological excavation material etc.).

These are just some of the main distinctions within different areas of research. Each distinction covers a wealth of subdivisions and differences in knowledge production ways and methodological approaches. To again use the social sciences as an example, qualitative studies cover methods such as individual and focus group interviews, participant observation, discourse analysis, and case studies, to name a few core methods, and quantitative studies can be anything from surveys to, for example, quantitative content analyses or hypothesis testing experiments. Such differences, which can also be found within other main areas of research, naturally also mean that there are important variations in the way in which GenAI is and can be used as well as differences in the potential and risks of using GenAI to support the research process.

It is impossible – within the scope of a single study like the present one – to map all the differences in knowledge production ways and the corresponding differences in the way in which GenAI is used, together with the associated potential and risks of using GenAI. Therefore, in this report we can only give inspirational examples from across disciplines and knowledge production ways. For this purpose, we will use an analytical model of the research process with five phases (cf.

Ravn and Sørensen, 2021; Andersen et al., 2024). Within each research phase, we will describe several cases of GenAI use and discuss potential and risks related to them. For comprehensibility's sake, we also present a table with additional examples of GenAI use in each research phase based on our literature review.

The analytical model consists of the following research phases:

a) Idea generation and research funding
b) Research design
c) Data collection
d) Data analysis
e) Scientific publishing, reporting and dissemination.

In the following, we will take a deeper look at these phases and describe how GenAI is being used to support the work conducted in the individual research phases.

4.2. Using GenAI for idea generation and research funding

The initial phase of research is centred around the process of generating ideas and writing research funding applications. This phase often includes literature reviewing, brainstorming sessions and discussions with other researchers, as well as an identification of gaps or emerging trends in current knowledge. Other tasks include the formulation of research questions and/or hypotheses. As these ideas take shape, researchers can utilize them for writing a research proposal or funding application.

The emergence of LLMs, with their capacity to process large amounts of information as well as generate new content, has significantly influenced this phase. LLMs introduce new practices to this phase where researchers already use tools like Elicit or ResearchRabbit to search for relevant literature and assist the literature review process (Whitfield and Hofmann, 2023; Cole and Boutet, 2023).

A case study from the aviation industry illustrates one way of using GenAI tools in the first research phase (Walton and Watkins, 2024). In the study, the researchers demonstrate how GenAI tools can be used to **generate insights, identify gaps, find themes, and propose directions for future research**. The case concerns the use of additive manufacturing (3D printing) which according to the authors has a major potential in the aviation industry, but a cautious approach has so far been taken due to flight safety.
Walton and Watkins (2024) describe how the researchers fed ChatGPT a previously published paper on the use of additive manufacturing in supporting the future aviation supply chain, then prompted ChatGPT to review the text, consider other published academic research within the field and identify gaps in the literature based on the paper's content. Additionally, the researchers also used ChatGPT to summarise and synthesise data collected manually, prompting it to outline the most important insights from this data. While the output from ChatGPT was general, the authors claim it nevertheless provided them with relevant insights into areas of needed research in their field.

The study exemplifies a simple way to use GenAI: feeding ChatGPT with a previous paper and prompting it to highlight gaps and potential new areas of research, based on that paper and the model’s own data. Other studies have used a similar approach to generate research questions. Agathokleous et al. (2024) prompted ChatGPT to produce 100 questions facing plant science, and Park et al. (2024) tested ChatGPT’s capability to generate hypotheses through basic interaction. Both studies find potential in the use of GenAI for generating ideas and hypotheses.

It is a method that can be applied across disciplines. This particular use of GenAI allows researchers to examine large amounts of data extremely fast, while revealing themes and patterns that might not be immediately apparent to researchers. Walton and Watkins (2024) argue that GenAI's ability to synthesise information into coherent narratives and insights is invaluable, because it improves the research process and enables researchers to identify critical areas of focus. A clear risk related to this case is, as the authors also note, that ChatGPT (at this point) does not identify the sources of content and bias cannot be determined with certainty. This issue is commonly referred to as "AI hallucinations", i.e., the notion that an AI appears to be generating false facts and sources. It is a risk which we will elaborate on in chapter 5.2.

Another study (Pride et al., 2023) approached this particular issue. They developed their own platform, CORE-GPT, to specifically address the problems of false and fabricated information from ChatGPT. Their platform is trained solely on a corpus of full texts of more than 32 million scientific articles, which are subsequently referenced and cited. Their results show that the platform, on the whole, provides comprehensive, useful and above all, trustworthy answers. For research purposes, GenAI models trained on fit-for-purpose datasets, could be a solution to the problem. However, the need for researchers to always interpret and validate GenAI outputs remains important when GenAI models are used.

According to Mads Rosendahl Thomsen from Aarhus University, we could be witnessing a monumental shift in the history of humanity. After millennia of doing things by hand, he believes a paradigm shift is now occurring in writing with the recent developments of LLMs. However, according to him, many researchers still think it is a bit embarrassing to admit that they are using GenAI tools to assist them with different writing tasks or idea generation. This is especially the case within his area (literature studies) because creating text here is a vital part of the research process. He uses ChatGPT and other GenAI tools himself for sparring ideas and for checking that he has covered all important points when writing about an issue. He adds that in most fields there are no good reasons for not using these efficient tools in the writing process, as long as the content is thoroughly checked by the author and there is transparency around the use of GenAI tools. The important issue is that authors are accountable for being able to explain their writing.

On the other hand, one of our interviewees, archaeologist and data scientist David Stott, expressed a general scepticism towards using GenAI for idea generation: "A lot of the time in our research we're interested in edge cases and outliers and we're trying to find the extraordinary, not the norm. And what the generative AI returns, generally, is a generalisation of a corpus of knowledge and that's not why I do what I do."

GenAI can also be used in the **writing of research (grant) applications**. Godwin et al. (2024) developed their own GenAI-based software that assists in writing grant application drafts. The task of writing a research application is complicated, time-consuming, requires administrative acumen and the competencies to articulate oneself succinct and precise. However, it is an essential part of the research ecosystem as most academic researchers rely on these applications to fund their research. The ability to navigate this task can make or break academic careers (Bol et al., 2018).

Their GenAI software functions as a support tool that helps researchers write application drafts through three functionalities: it searches previously funded applications and compares them to the proposed research; it assists drafting a specific aims page, highlighting the importance of the research; and it supports drafting key sections of the application, including the overall research strategy. The software utilises a database of previously awarded research applications within specific departments, enhancing the accuracy of the results. It also provides links to previous applications, thus addressing the issue of missing references.

Using GenAI for writing grant applications has several potential benefits. Apart from its obvious potential for efficiency gain, GenAI is also argued to contribute to "democratizing" the research funding process. This potential could be realised by supporting actors that traditionally might be at a competitive disadvantage, including junior scholars lacking the craft of "grantsmanship" (Godwin et al., 2024) and non-native English-speaking researchers (Hwang et al., 2023). However, this democratizing potential necessitates equal access to relevant tools, which is currently not (yet) the case.

A possible risk inherent to using GenAI for writing research funding applications (or for writing tasks in general) is the loss of diversity in the applications. This streamlining of writing applications could curb originality, as researchers may increasingly mimic each other using GenAI. Using GenAI in this process risks reducing the personal touch in an application and making it more superficial.

The grant application process does not stop at writing. The interviewee Johannes Bjerva from Aalborg University shared that LLMs can also help in preparing for grant interviews, which are part of some application processes (e.g. ERC funding). Using the new voice interaction feature of ChatGPT, he prompted the tool to ask questions (from his application, the evaluation he got, and general questions supplied by the fundraising team), answered them out loud, and had the model give him feedback. While the feedback aspect did not really work, he found the overall practice very helpful: “To just be in this mindset of getting the questions asked live and having the pressure to answer directly and not just having it on a piece of paper and reading it to myself. That was very useful.”

In relation to using GenAI for such purposes, Bjerva says that researchers must think about whether they mind that the application or other texts are used as training data for the GenAI model. There is a function in ChatGPT, for instance, to opt out of providing one’s data. For privacy and copyright reasons, he warns against uploading other people’s applications without the opt-out function switched on.

Table 4.2: Examples of use cases for idea generation and writing research applications

Research area	Use	Reference
All	Providing evidence-based answers to research-related questions	Pride et al., 2023
All	Using LLMs to aid non-native English-speaking researchers refining and editing text	Hwang et al., 2023
All	Using ChatGPT to generate scientific hypotheses	Park et al., 2024
Arts	Exploring interactions with AI-generated dance sequences as an inspiration source in dance composition and improvisation	Wallace et al., 2024
Health	Applying LLMs for article classification and categorizing large numbers of papers	Raja et al., 2024
Health	Using GenAI to help draft sections of grant applications	Godwin et al., 2024
Physical science	Using GenAI for information synthesis from various sources, not only academic papers	Zhao et al., 2024
Plant science	Using ChatGPT to generate research questions in plant science	Agathokleous et al., 2024
Production management	GenAI’s application in identifying research gaps	Walton and Watkins, 2004

4.3. Using GenAI for research design

The research design phase involves establishing the overall strategy and framework for the research project. The main tasks in this phase include defining research objectives, determining the approach, and selecting methods for data collection and analysis. This phase ensures that the project is systematically and logically planned, allowing for an accurate and reliable exploration of the research questions and hypotheses. Some of the work in this phase might take place in the idea generation and application phase, discussed above, but once funding is secured, a more detailed design of the project is often needed.

In our literature review we only found a few explicit examples of how GenAI can be used in the research design phase. Usually, the design phase was implicitly discussed together with other phases of the research process.

One notable example of this is the multi-LLMs-based intelligent agent, Coscientist, developed by Boiko et al. (2023) for chemical research. Coscientist integrates various relevant research tools and activities, such as internet and documentation searches, coding tools and robotic experimentation platforms. According to the authors, these features make Coscientist a versatile and powerful tool, capable of autonomous designing, planning and executing complex scientific experiments, all based on a simple plain text input (prompts) from a user. Coscientist, and similar GenAI tools, seems most relevant to disciplines where experiments and/or coding are essential components.

Similar to the other research phases, the main benefits of using GenAI for research design purposes seems to lie in its efficacy and time-saving prospects, as well as its versatility and comprehensiveness as a research tool. Another noteworthy feature of Coscientist, according to Boiko et al. (2023), is its capacity to provide justifications for specific choices, addressing concepts such as reactivity and selectivity. Coscientist's ability to address these questions might enhance the trustworthiness and utility of GenAI in research. A conspicuous risk, which the study also emphasizes, is the use of GenAI to conduct illicit activities. Just as Coscientist can be used for legitimate research, it can also be (mis)used and manipulated to design, analyse and plan the synthesis of illegal drugs and chemical weapons. This underscores the importance of implementing safeguards and ethical guidelines for proper GenAI use in research.

Bai et al. (2024) describe a related case, using six different open-source LLMs to basically conduct a full study; from knowledge extraction and research design, to experiment, data analysis, and paper polishing. The researchers argue that the LLMs have a high potential as powerful tools which can also assist in connecting all sections of the research process.

On a more general level, a study by Dashkevych and Portnov (2024) experiments with the use of GenAI for **enhancing their own research design**. Their study seeks to analyse how GenAI tools can assist in design and implementation of research within a particular research field: definition and identification of smart cities. The researchers compared responses and solutions generated by three different AI engines – ChatGPT, InferKit (now discontinued), and DeepAI – with findings from their own published research to assess whether these GenAI tools could improve the original studies. The comparisons were based on a series of questions to the tools, whose responses were subsequently evaluated using a list of pre-defined assessment criteria. All three AI tools displayed common shortcomings of GenAI products, e.g., incorrect answers, citing non-existent references, etc., but the researchers also argue that the tools can help streamline the research design by supplementing missing or overlooked information.

Another example of GenAI use in this research phase relates to **protocol writing,** mainly applicable to the life sciences. As the amount of data and literature keeps increasing, it is important that information is provided in protocols in such a way that machines can automatically find and use it. Protocols are, however, seldom reusable due to notable variations in how they are written, and the writing of protocols is a highly time-consuming task. Jiang et al. (2024) address these issues with their LLM, ProtoCode, which curates existing protocols from the literature and automates the generation of standardised protocols. They argue that LLMs can play a crucial role in enhancing research reproducibility through these automated protocol writing processes, and LLMs can at the same time relieve the work of writing these protocols.

One of our interviewees, Johannes Bjerva, explained that he had previously experimented with using GenAI tools for design purposes. His experience was, however, that it did not work well for this purpose and the output generated was far too generic. In general, our interviewees scarcely touched upon GenAI for research design purposes, which correlates well with the low number of case examples in the literature.

Table 4.3: Examples of use cases for research design

Research area	Use	Reference
All	A LLM tool that finds relevant experts, mainly as suggestions for collaborators	Zhang et al., 2020
Chemistry	Having LLMs do a comprehensive set of tasks, including knowledge extraction, database reading, experiment design and more	Bai et al., 2024
Chemistry	Utilizing a GenAI system that autonomously designs, plans and performs complex experiments	Boiko et al., 2023
Computational urban science	AI is used to create cities and urban design which can be used in weather and climate simulations	Aliaga and Niyogi, 2024
Economics	Using GenAI to uncover research objectives, scientific data, and models	Zheng et al., 2024
Environmental sciences	Using GenAI to assist in design and implementation of research on smart cities	Dashkevych and Portnov, 2024
Life science	Leveraging LLMs to create standardized research protocols	Jiang et al., 2024
Mechanical engineering	Using a GenAI tool to enhance coal mine research equipment	Cao et al., 2024

4.4. Using GenAI for data collection

The data collection phase refers to research practices aimed at obtaining *data* for answering research questions. Data are units abstracted from empirical phenomena of interest, selected from all the potential representations of a phenomenon for a certain purpose (Kitchin, 2014). Through processing, organising, and interpreting these data, researchers produce knowledge. The practices of doing so vary significantly between and within research disciplines and fields. Another name that can be used for this phase is data generation: using this term instead of 'collection' emphasises that data "do not pre-exist their generation; they are not simply waiting to be collected or harvested in a technical, passive and objective manner. Instead, they are produced – that is, actively created via procedures and instruments of our devising" (Kitchin, 2021, p.5).

For some research projects, this phase is dedicated to producing primary data: data "generated by a researcher and their instruments within a research design of their making" (Kitchin, 2014, p.7). Examples would be executing experiments and clinical trials, engaging respondents to answer survey or interview questions, setting up measurements through sensors, etc. However, much research relies on secondary data: data generated previously and for other purposes. These strategies are more like ideal types as many data collection processes combine those or can be put in between these categories: secondary data transformed into what will eventually be used for analysis can also be described as data "generated by a researcher and their instruments within a research design of their making" (ibid.). However, this dichotomy is analytically helpful for signifying variation in the data generation practices and the degree of control researchers have over the data.

One task that GenAI tools can be used for is **data retrieval**. Even if the research design does not entail generating primary data, creating usable data for analysis from existing sources might still be laborious. The information researchers need for their work is not always easy to find or access even if it is 'out there'; it can be scattered across multiple places and be embedded within a sea of irrelevant material. The literature we reviewed indicated two promising ways generative AI tools based on pre-trained models can be utilised here: to access and deliver information for the research dataset. For example, ChatGPT as well as multiple tools based on it facilitate interacting with PDFs (a format quite restrictive in terms of interaction) in an LLM way, by posing questions and receiving answers (Formanek, 2024). This can save time, especially when a tool allows question-answering to be conducted in a language different than that of the PDF, making working with documents in a foreign language easier. However, the quality and trustworthiness of the answers might be suboptimal, which means this process requires checks. Other use case examples would be extraction of metainformation from scientific literature for reviews and bibliometric research (Guo, M. et al., 2023), scraping the web and interacting with databases.

GenAI tools can also identify relevant information within different sources and retrieve only that relevant information. To use pre-trained models like LLMs for specific tasks, researchers must give the models instructions, which can be more or less elaborate. For example, Hu et al. (2023) propose a method for training GenAI tools to recognise location descriptions from tweets by providing models with 22 examples of tweets and parts from the tweets labelled as location descriptions. The tools then capture the location descriptions from a larger dataset of tweets. This result can be achieved in other ways, such as supervised machine learning methods for natural language processing (Stock et al., 2022). However, those require larger training datasets (Hu et al., 2023), which require more time and annotation work. The benefit of pre-trained models is that they can be relatively easily infused with new patterns (Hu et al., 2023) and require less technical expertise from the researcher

(Wang, S. et al., 2024). This use case has potential for projects across research areas that need to comb through many sources to compile their datasets. The main concerns with this use case would be about the quality of the produced samples and the use of resources (time, labour, energy, etc.) compared to other methods. Those considerations would vary based on the precision needed for the task, the volume of the material, availability of existing tools that fit the task well, etc.

Moreover, data might need **conversion** to be fit for the analysis researchers plan to execute. For instance, researchers usually generate transcripts from interview recordings to work with textual data for analysis. Automated speech recognition tools, such as software based on the OpenAI's Whisper model, can assist with this laborious task. Wollin-Giering et al. (2024) review multiple automated transcription tools and conclude that automated transcription tools save time compared to initial transcription, but the time-saving benefits would vary between tools, local hardware setups, and factors that affect the quality of the automated transcript and thus the time for correction (audio quality, language spoken, speech tempo, etc.). The authors also highlight that researchers must be mindful of privacy and regulatory considerations (where data is stored and with whom it would be shared) regarding different tools when selecting which one to use.

However, the most significant application of GenAI in this phase is generating **synthetic data**. As mentioned before, generative models can produce new data samples based on the probabilistic distribution of the training data. This allows **synthetic data augmentation** for computational research. Generating new data samples can alter data in ways that would improve the robustness of the model. For example, 'real' images can be rotated or segmented. While augmentation methods exist outside of GenAI, this is an area where generative models have a huge impact on the research process. According to one of our interviewees, Jes Frellsen from the Technical University of Denmark (DTU), GenAI methods are increasingly replacing previously existing methods for tasks like data augmentation: "GenAI is basically replacing human handcrafted rules and models, similar to how it's replacing human text editing with transformers."

Generative models can also be used to increase data volume by generating more samples. Many machine learning methods require (a lot of) training data to perform well, and data scarcity is a problem for research, e.g., in the medical domain (Sindhura et al., 2024). To mitigate this problem, synthetic data is generated to increase the volume of training data so the resulting model will perform better. For example, Chui et al. (2020) enhance a dataset on students' academic performance to improve prediction accuracy. Synthetic data can also be used to substitute for missing data, although this imputation approach might not always be the best course of action (Muhammad et al., 2023). This can be beneficial for many research areas that utilize machine learning for classification/detection and suffer from data scarcity.

Researchers can also enhance data of particular categories within their dataset to correct data for training models. Imbalanced training datasets can affect the models' performance: for example, if there are very few examples of abnormalities in the data, the model would predict these abnormalities with lower accuracy than if the classes were represented more equally. To achieve that balance, generative models such as GANs can be used to produce more data samples for the underrepresented classes. For instance, researchers in the plant disease field have been working on various generative approaches for diversifying plant image datasets; such augmentation of data improves the accuracy of plant disease detection based on images (Muhammad et al., 2023). In that way, generative models can support research done with other methods, such as classification models. However, this approach still needs enough minority samples, otherwise the improvement will be less significant (Liu et al., 2022).

Augmenting the data with GenAI comes with some risks for the research results. We discussed data augmentation in our interviews with, among others, Jes Frellsen from DTU. He notes that introducing synthetic data affects the algorithm trained on the data. Overall, he emphasises that it is crucial to understand the augmentation process and check that it makes sense for the task and produces realistic outputs.

Jes Frellsen also points out that there is a difference between modifying the existing research data and generating fully synthetic data with generative tools, as in the latter case the risk of producing data that are too different from the rest of the training dataset is much higher.

Researchers have also experimented with working with synthetic data only (although any kind of synthetic data produced by generative models has some connection to some data that was used for training). The practice of using **simulations**, as in all aforementioned use cases of generative AI for data collection, predates the current GenAI technology. For example, it is commonly used in epidemiological research, and here GenAI has the potential to improve simulation methods (Wang, H. et al., 2022). However, the release of LLMs also prompted tests of whether these models can be used as simulations of traditional 'human subjects' data methods such as interviews, surveys, etc. The idea is that LLMs' responses would align with the 'real' human narratives due to their knowledge base coming from large training data.

The reported results within our literature review are mixed. Cappelli et al. (2024) found that LLM responses to a previously unpublished questionnaire were close to those of expert responders. However, Bojic et al. (2024) observed that GenAI predictions did not compare well with field data on pain experienced by rock climbers and included incorrect information (such as fake references). Thus, Bojic et al. (2024) argue that publicly available GenAI tools do not perform well enough (yet?) to be used in pain research. Potentially, more domain-specific models trained on pain data would perform better. Dengel et al. (2023) tested an interview simulation with LLMs instead of human respondents and concluded that the conversations achieved were 'human-like' but not particularly profound. For example, they noted that LLM responses tended to be generic and not specific to the specific topic of conversation. Moreover, they argued that at least for their topic (computer science education in schools) the knowledge/ infor-mation element in the answers did not come close to expertise. Human-computer interaction scholars Hämäläinen et al. (2023) also generated data through LLMs and tested the LLM answers against human and concluded that results are human-like but different from what a qualitative study would obtain (for instance, synthetic data is less diverse). They also argue that if one is really interested in what people think or feel, one has to ask the real people. Indeed, this idea of creating artificial human subjects, regardless of the quality of the generated data, might go against knowledge practices within more interpretative epistemic cultures and disciplines, where synthetic data might be seen as lacking validity and even being close to data fabrication.

However, Hämäläinen et al. (2023) also point out a less 'synthetic' use case for GenAI in qualitative research: synthetic data can support research design development or piloting as its form resembles 'real' data. This last use case again emphasises how intertwined the research phases are, as indeed one can collect data for a pilot study to develop the research design of a bigger project.

Overall, this chapter shows that GenAI methods have already been used for a variety of data collection/generation tasks and have a lot of potential there. Using generative models here can save time, costs, and even improve the research results (for instance, develop better classifier models). Some of these benefits concern more computational research, while some lower the barriers for researchers with less technical/coding expertise (for instance, delegating data retrieval to LLMs instead of performing NLP methods).

Another potential for the researchers is that synthetic data can be seen as a solution to regulatory, ethical, and logistical issues around data collection and (re)use, which contribute to data scarcity issues, for instance, in medical and healthcare data. Adam Hulman from the Steno Diabetes Center in Aarhus, who was interviewed for this study, sees a future for synthetic data in his field as "data sharing in the healthcare domain is just really painful, sometimes even between collaborators or between the region and the university." He points out that epidemiological studies are expensive, and the resulting data are not used enough because they cannot be shared or because the sharing procedure is not open. Thus, he expects that synthetic health data in the future can contribute to better (re)use of health data and the promotion of the FAIR principles, i.e. that data should be Findable, Accessible, Interoperable, and Reusable.

In this sense, synthetic data has the potential to both lower the burden of data production for society as well as simplify data access for researchers. However, synthetic data can also pose some ethical and legal concerns, such as questions on whether it really safeguards against deanonymization or identification (Giuffrè and Shung, 2023). It is a burgeoning field of study, with researchers actively searching for both technical and ethical-legal solutions to maximize the potential of synthetic data for research as well as minimize risks and misuse.

Table 4.4: Examples of use cases for data collection

Research area	Use	Reference
Archaeology	Using GANs for producing multidimensional synthetic data which can help augment datasets for greater predictive visualization	Courtenay and González-Aguilera, 2020
Archival and library studies	Using fine-tuned LLMs to automate archival research, including document retrieval, document summarization, and rule-based compilation	Guo, D. et al., 2023
Cardiology	Enhancing data of a few categories of electrocardiogram signals using an improved GAN model	Liu et al., 2022
Computer sciences	Improving the performance of numerical weather prediction models with the use of GANs	Jeong and Yi, 2023
Computer sciences	Using GenAI to make realistic face photo-sketch synthesis	Yu et al., 2021
Engineering	Using GAN to research and improve automatic textile defect detection methods.	Rui and Qiang, 2022
Plant sciences	Using GenAI for plant disease image augmentation, addressing issues related to data scarcity	Muhammad et al., 2023
Qualitative social sciences	Using LLMs as interviewees, simulating opinions of a specifically targeted audience	Dengel et al., 2023
Technical and computer sciences	Generating multiple code explanation types using LLMs	MacNeil et al., 2023

4.5. Using GenAI for data analysis

A core phase in any research process is the data analysis phase. Data do not speak for themselves, they must be analysed and interpreted by researchers to make scientific sense. This is the case no matter which area of research one works in, and many different strategies and methods are employed to analyse and interpret data in this phase, ranging from quantitative approaches and methods such as statistical analyses and mathematical modelling, to qualitative methods like coding and analysing interview material or interpretating art and literature.

Given the vast differences that exist in the research system between disciplines, epistemic cultures, and approaches and methods, the benefits of using GenAI to assist the work in this phase will similarly vary greatly from research field to research field. Nevertheless, our literature review shows that researchers from many different fields are already applying or have started experimenting with GenAI tools and models in the data analysis phase.

GenAI is, for example, used for analysing human behaviour, such as predicting pedestrian movement (Pang et al., 2022), or to assist in mental health research by identifying risk factors related to suicide (Lissak et al., 2024). Different forms of GenAI have likewise been applied to create and reconstruct cultural artifacts, such as Mongolian ethnic patterns (Wu et al., 2023), or to address cyber threats by simulating ransomware behaviour, enabling better preparedness and response strategies (Urooj et al., 2024). Radiology researchers have also experimented with using a GPT plug-in for data analysis and visualization, although noting that this GenAI approach was very speedy but (at the time) produced mistakes and required verification (Bhayana, 2024).

Despite the many examples of GenAI use in the data analysis phase described in the literature (see also Table 4.5), it is important to note that there are substantial differences in how relevant it is to use GenAI for analytical purposes across research fields and knowledge production ways. In some fields there seems to be very little or no added value of using GenAI for analytical purposes whereas GenAI appears to be revolutionizing other fields. In our interviews, we came across both ends of this continuum. David Stott, archaeologist and data scientist working at Moesgaard Museum, uses non-generative AI for automated feature detection of very large datasets, e.g. stemming from satellite images. He also works on national or subcontinental scale LiDAR (Light Detection and Ranging) data sets with up to 20 terabytes of data and use them for searching for topographic landforms, looking for things like ring fortresses or other features interesting for archaeology. To him, there would be no added value in using GenAI for such analytical tasks. In general, he is also sceptical towards using GenAI in archaeology apart from in relation to more mundane tasks like writing code or translating language. It could, he says, be used, for example, to create images of what a village might have looked like in the Viking Age, or to interpolate fragments of artifacts into complete artifacts. In both cases, however, he points out that it would be difficult to validate such findings.

At the opposite end of the continuum, we find Timothy Jenkins from DTU, who was likewise interviewed for this report. Jenkins conducts research on issues related to biotech and biotherapeutics. He especially focuses on protein structures and combines both computational and laboratory-based approaches. Jenkins uses the GenAI tool RFdiffusion, developed by Professor David Baker from the Department of Biochemistry at the University of Washington, to design molecules and optimize therapeutic compounds that are then later tested in the lab (cf. e.g. Watson et al. 2023). According to Jenkins, RFdiffusion is "… the biggest advancement that my field has ever seen. It has really transformed our ability to design therapeutics computationally." Using RFdiffusion in combination with the sequence-design tool

ProteinMPNN and AlphaFold that is used for predicting the folded shape of proteins, has according to Jenkins made it possible to increase the success rate of the designs working in the lab from 0.1% to 30-40%. The success rate might even increase more according to Jenkins, who says that "… now we've found ways to tweak it even further, where we're getting closer to 80% success rates. So, it has really transformed my area of research. I think it's a very small percentage of people who have realized how transformative this technology is." He adds that it will change the way therapeutic drug discovery works, because the time it takes to develop drugs and get them into the market will decrease heavily. What used to take up to a year to do can now be done in a month's time by using these new tools: "… and you could probably do it even faster."

Together with increased speed and efficiency, Jenkins points to saving costs as one of the major advantages of using GenAI in drug development. It is much cheaper to test hypotheses and do experiments on the computer than in the laboratory, and this again opens for developing drugs for rare diseases and diseases that primarily affect impoverished communities in the Global South (like snake bites), which so far have been underexplored due to lack of funding and lack of commercial interest. Jenkins mentions that a possible risk of developing and using tools like RFdiffusion in combination with the other AI tools, is that this technology could also be used with malicious intents, e.g. for trying to develop bioweapons. However, to produce such weapons, one would also need substantial know-how, a well-equipped laboratory, as well as a way of circumventing existing safeguards. Like many of the other interviewees, he also points out that we must be aware of the possible environmental costs of using these technologies due to increased energy consumption.

Drug discovery is undoubtedly among the most interesting and successful cases of deploying GenAI in the data analysis phase and research process in general. There is a growing literature describing advances in this area (cf. recent review articles: Romanelli et al. 2024 and Tang et al. 2024) and it is interesting to see how this development is also deeply interconnected with private tech companies, like Google DeepMind that runs AlphaFold, mentioned above, and that recently (September 5, 2024) introduced AlphaProteo – a new AI tool for designing "novel, high-strength protein binders to serve as building blocks for biological and health research" – which according to the company will accelerate the discovery of new drugs even more (DeepMind, 2024).

Generative models seem to be particularly useful in **'design' research approaches**, where the goal is to create something new (an algorithm, structure, chemical, formula, etc.). In this type of research, one designs or creates a potential, hypothetical approach (algorithm, protein, etc.) and then validates it to see how it performs. In these studies, GenAI plays a role that can be positioned in-between or across three research stages: research design (formulating a hypothesis), data generation, and data analysis. It seems like GenAI is particularly helpful in these kinds of approaches, because it can help speed up the process and test different hypotheses, thereby improving the scope and scale of designs proposed and tested.

An example of GenAI's success in this kind of research design from another discipline – physics – was mentioned by the interviewee Jes Frellsen. He explains that generative models are highly effective at characterizing magnetic fields. While physicists already have robust methods for simulating these fields, generative models can produce predictions much faster at increased resolution, potentially allowing for more experimentation in optimizing magnets for specific applications.

The interviewee Simon Rasmussen from the Novo Nordisk Foundation Center for Basic Metabolic Protein Research at the University of Copenhagen also works with generative models. For example, he uses variational autoencoders to analyse extensive patient data across multiple datasets. He notes the impact generative models have made in medical research. Generative models, for instance, can be manipulated to get insight into the latent representations they learn, from which researchers can shed light on predictive factors of the phenomenon of interest (e.g., risks of disease or drug efficacy). However, Rasmussen notes that in this area of study generative models have been less transformative than in the protein and drug design area. But while specialized LLMs, for instance, have been a game changer in designing proteins or drugs, he points out that it is important to understand what these models can and cannot do, as “they will not solve everything within human biology.” Further, he notes that the inherent bias that comes with working with medical data, which reflects not only the patient’s biology but also the way a medical system works, makes it trickier to utilise generative models: it can be difficult to not mix up the biology with the doctors’ behaviour.

Thus, the impact of GenAI across areas of research depends on the type of research problem, the nature of the data an area of research draws on, and probably other factors as well. While it can be truly revolutionary for certain research fields, it is not a solution for every research problem.

Table 4.5: Examples of use cases for data analysis

Research area	Use	Reference
Biology/animal sciences	Applying a GAN to perform image-to-image processing steps much faster than manual processing	Robson et al., 2021
Cognitive computation	Using a GenAI based model to detect potential cyber attacks	Islam et al., 2024
Computer science	Colorization of Chinese farmer paintings based on GANs	Peng et al., 2023
Health	Using LLMs to address current rehabilitation issues, including data bias, contextual comprehension, and ethical concerns	Bonnechère, 2024
Nuclear power	Detecting anomalies in the operation of various technical systems using deep generative models	Li, X. et al., 2022
Ocean sciences	Predicting levels of dissolved oxygen concentration in sea water	Chen et al., 2024
Orthopaedics	The application of LLMs and ChatGPT to support various diagnostic tasks in orthopaedic disciplines	Chatterjee et al., 2023
Political science	Automated annotation of data (e.g. tweets, posts, images) for systemic analysis	Linegar et al., 2023
Qualitative research (in software engineering)	Analysis of textual data stemming from qualitative research, including the identification of themes, pattern recognition, etc.	Bano et al., 2024
Scientometrics	Using LLMs to perform automated taxonomy alignment which can help bridge gaps between knowledge domains	Cui et al., 2024

4.6. Using GenAI for scientific publishing, reporting and dissemination

The final phase of the research process encompasses the transformation of research findings into knowledge that is accessible and useful to various audiences. In some disciplines, this occurs after data has been analysed and processed, while in others it may be more integrated in the data analysis phase, e.g. in some parts of the humanities (Laudel, 2024). This stage includes tasks such as writing research articles or reports; presenting work at conferences, symposia and other academic meetings; and communicating research findings to non-academic audiences, for example through press releases, blogs, podcasts and the like. We also include in this phase tasks that involve evaluating the work of others, usually through peer review for academic journals. The use cases that we describe in this chapter are all general, in the sense that they relate to no academic disciplines or knowledge production way in particular, but instead are relevant across the board.

It is clear that GenAI has potential to considerably affect this phase of the research process, encompassing some of the more obvious use cases for GenAI. After the introduction of LLM-based models such as ChatGPT, the initial attention of the wider research community was focused on their applications in writing research articles or other textual content such as conference abstracts.

Studies indicate a significant share of researchers using these tools for **academic publishing** purposes, particularly to generate full research papers or refine existing text for publication. While detection of GenAI usage still faces considerable challenges, studies using various methods, usually based on changes in writing style or vocabulary in large corpora of texts, find estimates of GenAI-affected language in about 1% to as much as over 30% of articles published in 2023 (e.g. Gray, 2024; Kobak et al., 2024). All of these studies are probably conservative estimates. Similarly, studies report significant use of GenAI tools in peer review reports (e.g. Latona et al., 2024). While the benefits in terms of productivity and efficiency are apparent, the use of such tools in peer review is prohibited at almost all major publishers, journals and funding agencies, currently being one of the few GenAI use cases having generally implemented regulations. The main argument for prohibiting this use relates to privacy, confidentiality and intellectual property right concerns. In addition, the accuracy and reproducibility of GenAI tools for research assessment purposes remains uncertain (Thelwall, 2024; Latona et al., 2024).

While general purpose LLM-based tools such as ChatGPT can assist with academic writing, some more specialised tools have been developed, including AUTOGEN (Mann et al., 2023). This tool takes an author's own writing as input and aims to produce new text that mimics the author's style. In addition to the benefits of general-purpose LLMs, such as productivity and efficiency gains, this tool arguably has the advantage of preserving writing styles and cultural traditions. However, it creates a risk of reducing diversity and raises concerns about plagiarism, privacy and intellectual property rights. Furthermore, while GenAI tools promise to level the playing field for researchers with language barriers or limited academic writing experience, tools like AUTOGEN may offer limited support to these individuals. Overall, the tool has been critically reviewed by various scholars (Resnik and Hosseini, 2023; Erler, 2023). In a further attempt to develop more specialised tools for scholarly publishing tasks, some of the larger commercial publishers are developing in-house peer review tools, thereby aiming to circumvent the concern over confidentiality and privacy issues when using third-party software.

One of our interviewees, Daniel Russo from Aalborg University, sketched an important potential risk associated with GenAI-assisted publishing. Research evaluation and assessment usually rely heavily on the number and outlets of research articles published by an individual or institution. For many fields, journal articles are the main unit of assessment, but in some fields like computer science and software engineering, conferences and their proceedings are highly valued. GenAI challenges the traditional conference and publishing model because it has made it much easier to produce content for academic papers and to write papers (using some of the tools mentioned above). This has resulted in further acceleration of the number of papers written and submitted to journals and conferences, leading to barely sustainable extra work for reviewers, and challenges for junior researchers, who often have to get papers accepted to get a tenured position. Russo argues in favour of rethinking conferences and publishing models, i.e. to focus more on the things that are unique to humans, like being creative, thinking and working together to solve problems, focusing more on originality rather than mainstream solutions.

GenAI can also be used on 'the other side' of the publication process, for peer review and editorial purposes. Partly because of the inflation in submissions, in which GenAI has played a role, an increased use of generative models for evaluating papers has been proposed. The interviewee Johannes Bjerva points out the difference between asking an LLM for support in *writing* the review report versus feeding the manuscript to the model and producing a review without having to read the text. While the former can be fine in terms of the resulting review (though intellectual property concerns might remain), the latter, argues Bjerva, is highly problematic as it would introduce more bias into the peer review process: and this is a process already prone to bias. If we consider that some publications can already be almost completely AI-generated, automated reviews become especially problematic (cf. Hosseini and Horbach, 2023).

Beyond academic writing and peer review, GenAI is being explored for enhancing **science communication,** i.e. communicating research results to non-academic audiences. Tools like Metaphorian (Kim et al., 2023), which suggests metaphors, and ZINify (Shriram and Sreekala, 2023), which creates science-based graphic novels (zines), aim to boost creativity and support researchers in areas outside their core expertise. Additionally, general-purpose LLMs have been used for summarizing research abstracts into lay language (Shyr et al., 2024), and specific LLMs have been trained to make clinical knowledge more accessible to patients (Lozano et al. 2024).

All these attempts have the common purpose of making research findings more accessible to wider audiences and could be used either by researchers wanting to disseminate their work or by audience members looking for information on a certain topic. On a more general level, this is part of an attempt to leverage GenAI tools to contribute to the open science movement (see Hosseini et al., 2024 for a comprehensive overview of the potential and potential drawbacks involved in this).

Third, GenAI has been used to translate research findings into **educational material**. The generation of text-based material for readings or lectures is a straightforward example, as might be applications of GenAI as a 'teaching assistant' or sparring partner for students to develop specific research skills. However, researchers have also experimented with using GenAI tools for the creation of educational games (Moon et al., 2024) or the design of virtual reality applications to improve understanding and application of scientific results (Schmidt et al. 2024). This potentially allows such material to be produced on a large scale and perhaps even personalised to individual users, whereas their production using traditional methods is very resource intensive.

Table 4.6: Examples of use cases for scientific publishing, reporting and dissemination		
Research area	**Use**	**Reference**
All	Having ChatGPT assist the draft of a scientific article	Altmäe et al., 2023
All	Utilizing GenAI to enhance English presentation drafts, skills and general pronunciation practice	Hirosawa and Shimizu, 2024
All	Using LLMs to assist the process of writing and revising scholarly manuscripts	Pividori and Greene, 2024
All	Having LLMs and text-to-image-based tools transform research papers into zines (i.e., short, magazine-like booklets)	Shriram and Sreekala, 2023
All	Using an LLM to enhance academic prose based on a specific author's previous writing	Mann et al., 2023
All	Using an LLM-based tool to create metaphors for science writing	Kim et al., 2023
All	Assessing whether ChatGPT is accurate enough to perform research evaluations on journal articles	Thelwall, 2024
Ecology	GenAI text generating tools are used to assist non-native English-speaking researcher disseminating their research	Zenni and Andrew, 2023
Educational studies	Using various GenAI tools for educational game design, allowing personalized designs for individual users, targeting wider audiences	Moon et al., 2024

5. Potential and risks of using GenAI – an overview

As we have seen in Chapter 4, GenAI has the potential to significantly enhance the research process – at least in some fields – by shortening the time required to perform specific tasks, enabling new discoveries, and providing a deeper understanding of the phenomena being studied. However, the use of GenAI in research also raises concerns, including issues like hallucination, ethical dilemmas regarding data privacy and biases, and research integrity-related concerns about reproducibility and accountability. In this chapter, we provide an overview of the potential and benefits of using GenAI in the research process (Chapter 5.1) as well as the main risks and concerns (Chapter 5.2), as described in the literature and articulated by our interviewees. Note that we do not aim to provide an exhaustive list of all benefits and concerns, but rather to describe those occurring most frequently in debates or those potentially having the most significant impact.

5.1. Potential and benefits

We divide the potential and benefits of using GenAI in the research process into three categories: **Scientific, Economic,** and **Societal** potential. In describing the potential, we use examples from Chapter 4 as well as additional examples identified in the literature.

Scientific potential

In a recent report, "Science in the age of AI", the Royal Society (2024) summarizes the benefits and potential of GenAI for research in this way:

"Generative AI tools can assist the advancement of scientific research. They hold promise for expediting routine scientific tasks, such as processing unstructured data, solving complex coding challenges, or supporting the multilingual translation of academic articles. In addition, there may be a place for text-generation models to be used for academic and non-academic written tasks, with potential implications for scholarly communications and research assessment." (The Royal Society, 2024, p.7)

In our study, we also found that GenAI has the potential to positively impact most research fields. However, as shown in chapter 4, there are substantial differences in how useful GenAI is for different fields, especially when it comes to data collection/generation and data analysis.

The drug discovery field, as discussed above, is an example of an area where GenAI is already profoundly transforming the research process. Its impact and potential may be less pronounced in other fields, but GenAI still has the potential to significantly enhance research across various fields by increasing both **speed and productivity**. For instance, it has already positively affected fields that rely heavily on coding. GenAI tools can assist researchers in generating code based on natural language descriptions, speeding up the debugging process by identifying errors or potential bugs, and completing coding tasks by predicting what developers want to write next. Besides speed and productivity, GenAI also seems to have the potential to **enhance research insights**, at least within fields where GenAI can be applied to analyse huge and/or very heterogeneous datasets. Here, GenAI might help identify patterns in the data that would otherwise be hard to discover.

If we look at the idea generation and research funding phase and the scientific publishing, reporting and dissemination phase, it seems that most fields can benefit from using GenAI (LLMs) to assist with at least some of the tasks within these research phases. This could, for example, be for **translation** purposes or for different **writing tasks**. Especially non-English speakers can get valuable help from LLMs for various writing tasks. We know that non-native English speakers use much more time on, for example, writing papers and preparing presentations in English, than native English speakers do (Amano et al. 2023). GenAI might not be able to '**level the playing field**' completely, but can make it easier for non-native speakers to create abstracts for conferences, write research applications in English, etc. (Hwang et al. 2023).

Additionally, it is important to note that research tasks are rarely the only duties researchers have in a workday. Typically, researchers must juggle various administrative responsibilities and teaching obligations alongside their research tasks. GenAI might be able to assist with some of these duties, potentially **reducing the burden of administrative tasks** (see, for example, Rengers et al. 2024) and thereby freeing up valuable time for research.

Economic potential

The economic potential and benefits of GenAI are closely connected to the scientific benefits. If GenAI can help speed up the research process by, for example, reducing the time it takes to perform a literature review, an analysis or a writing task, it has at the same time, the potential to **reduce the costs** of a particular research project and **increase the productivity** of the research system. Quicker research cycles will potentially also lead to **faster product developments** and **more commercialisation** of research in fields where this is relevant, providing research systems that are able to take full advantage of GenAI in the research process with a **competitive advantage**.

Societal potential

GenAI has already demonstrated its potential to revolutionise research areas with direct societal implications, including drug discovery, medical diagnostics, and climate science, as discussed in section 4.5. In fact, some of GenAI's major contributions have thus far been in translational and application-oriented research. In terms of the drug discovery process, GenAI has accelerated the process and made it more efficient. GenAI makes it possible to analyse biological data faster and predict interactions. In this way, GenAI can also help identify potential drug candidates more quickly for **rare diseases** that often lack treatment options, as well as for **diseases affecting people in impoverished countries**, which typically attract little commercial interest. GenAI – together with AI in general – might also make it possible to **address global challenges** like climate change, floodings (Wang, S. et al., 2024) or health crises more effectively through the increased opportunities for working faster and more efficiently, and the new possibilities it offers for analysing large and heterogeneous data sets. In addition, GenAI's potential to allow meaningful access to research outputs to wider, non-academic communities, could improve knowledge circulation and uptake within society (Hosseini et al., 2024).

5.2. Risks and concerns

In the following section, we will provide an overview of the main risks and concerns associated with using GenAI in the research process. Like in the previous subsection, we will list examples identified earlier as well as additional examples from the literature. We will categorise the risks and concerns into four groups: **Scientific**, **Economic**, **Societal**, and **Environmental**.

Scientific risks

When using GenAI for research purposes, a key concern is that GenAI models might generate outputs that contain information that is incorrect or misleading, despite sounding plausible and convincing. This could be incorrect summaries of material, wrong code, or made-up literature references. This phenomenon is often referred to as **hallucination** or simply lying. However, according to Hicks et al. (2024), it is important to understand that LLMs like ChatGPT are not directly concerned with the truth – with giving correct answers to questions – but designed to provide text answers (or sound, images etc.) to a question (prompt), no matter what. Instead of referring to incorrect or misleading answers as hallucination or lying, Hicks et al. (2024) therefore suggest calling the output of LLMs 'bullshit', because of LLMs' lack of relationship with truth.

Closely connected to this concern is the risk of **biases** in AI-generated content. GenAI models are trained on large datasets that may contain biases. These biases can stem from societal prejudices – for example, particular cultural representations of certain demographics – which can lead to skewed outputs and reinforce existing stereotypes (e.g., that a medical doctor is a man and that a nurse is a woman). Biases can also be the result of the underrepresentation of certain groups in the training data, which might result in the GenAI model performing poorly for these groups. This is especially problematic when GenAI tools are used for diagnostic purposes, for giving health and medical advice, and for societal interventions. There can also be algorithmic biases, i.e. biases stemming from how an algorithm processes and interprets data.

In the literature, **de-skilling** due to automation of research tasks (for example in relation to coding or data analysis) is also discussed as a possible risk connected to GenAI (see e.g. The Royal Society, p.26). There are also worries about **lack of originality** and heterogeneity in research output generated with the help of GenAI tools (see e.g. Grimaldi and Ehrler, 2023). With the hype surrounding GenAI, one might also fear that research and funding schemes become more focused on projects that can be done with the help of GenAI tools and that non-GenAI projects become less interesting for researchers and funders, even though they still might be relevant and important. The Royal Society expresses similar concerns, when they write that "… the changing incentives across the scientific ecosystem may be increasing pressure on researchers to incorporate advanced AI techniques at the neglect of more conventional methodologies, or to be 'good at AI' rather than 'good at science'" (The Royal Society, p.5). As discussed above, GenAI also makes it much easier to produce new papers in some fields, which poses new questions to the merit and career system within these fields and make it even more difficult to compare productivity and performance across fields.

In addition to these problems, it is also worth mentioning that there might be potential **research integrity** problems involved in using GenAI. As discussed above, GenAI makes it much easier to produce new papers in some fields, which poses new questions to the merit and career system within these fields and academia in general. Other integrity concerns include **reproducibility challenges.** These emerge both because systems develop very rapidly, making it difficult to track how research done with previous versions of a system corresponds to state-of-the-art models, as well as because GenAI systems have some degree of fundamental randomness build into them. The latter means systems will necessarily produce different responses to similar prompts or inputs, fundamentally complicating reproducibility or replication attempts. **Black-boxing** of the generative process – especially in commercial models – also limits explainability and makes it difficult to trace how input is transformed into output in these models. This can also make it difficult to verify results and can constitute barriers to the effective adoption of open science principles, according to The Royal Society (2024, p.5, 40-41) (see also Hosseini et al., 2024).

Economic risks

The main economic risk associated with using GenAI in the research process relates to **copyright infringements** when materials – such as texts, photos, and music – are uploaded and used as training data for GenAI models. This can have serious financial consequences for artists, writers, and private companies. GenAI models are dependent on data for training, making it crucial for both private companies and publicly (and privately) funded research projects to establish fair solutions and compensation schemes when using materials created by others. This issue has recently gained attention, particularly in relation to private publishers selling research articles to companies for training purposes without notifying or **compensating** the authors (see, e.g., Potter, 2024). Concerns over copyright infringements are also expressed in relation to the use of GenAI for peer review purposes and other occasions where third-party created content is uploaded to GenAI tools. Additionally, another more general concern regarding the economic risks is related to the **infrastructural costs** and the potential pitfalls of making poor investments – specifically, investing heavily in GenAI initiatives that may ultimately prove less successful than expected, thereby diverting funds from potentially more beneficial projects.

Societal risks

As discussed above, biases are one of the key concerns in relation to the use of GenAI tools in research. Any use of biased GenAI models raises ethical questions about **fairness** and **potential harm** to marginalized groups. Similarly, it is important to be aware of issues around **confidentiality** and **privacy rights** in relation to the use of these tools. In qualitative research, for example, researchers cannot just upload interview transcripts and research notes to openly available GenAI tools to make summaries or automated coding of the material. In Europe, researchers also have to be aware of **GDPR regulations**, making sure that personal and sensitive data is not involuntarily shared with private companies and the public via these GenAI research tools. If researchers want to use GenAI tools in the research process, it is important to get research participants' and Ethical committees' approval of this procedure, for example, by making it a part of the **informed consent** procedure and Ethics approval applications.

Further, it is important to recognize the role that **private companies** play in the rise and continuous development of GenAI, and the vulnerability and dependency this creates for the research system. Private companies control access to the tools and services they have developed – and they determine how much is charged to use them. Moreover, in recent years the industry has become more and more involved in the field of AI research (Klinger et al., 2022), increasingly **impacting which technologies** and tools are developed. When making decisions about future investments and which technologies to focus on, private companies have their shareholders' interest in mind. Their decisions will not always correspond to the interest of the public and the research system. Any dependency of private companies' GenAI tools and services, therefore, constitutes a risk for researchers, research projects, and research systems. Finally, there is also the general fairness question of **who benefits** from GenAI use in research. There is no doubt that researchers in affluent countries and institutions can benefit greatly from these technologies, but the same is not always the case for researchers working in less affluent societies and institutions, for example, in the Global South. They might face problems affording access to these tools, which again can increase existing **inequalities** in the global research system and between societies. Lastly, the effort of low-paid workers from marginalized communities, who for 1-2$ per hour perform parts of the essential training tasks that enable GenAI systems to learn and evolve, should also be mentioned and recognized (Rowe, 2023).

Environmental risks

The biggest environmental concern related to GenAI is the **carbon footprint** and **increased electricity consumption** connected to this technology. A ChatGPT request is estimated to require around 10 times as much energy as a Google query (Coskun, 2024). With the success of GenAI, there is an increasing demand for more data centres. Currently, there are about 8,000 data centres worldwide, with 1,240 located in the EU (2022 figures, cf. International Energy Agency 2024, p.32). In 2022, data centres accounted for approximately 4% of the total electricity consumption in the EU, a figure that could double by 2026 due to the additional energy required for AI (and cryptocurrencies). In Ireland, which is a major hub for data centres, these facilities consumed 17%

of the country's total electricity in 2022, a number projected to rise to 32% by 2026 (International Energy Agency 2024, p.32). Similarly, in Denmark, where there are also a high number of data centres, their electricity consumption is expected to reach around 20% of the country's total electricity consumption by 2026 (International Energy Agency 2024, p.34).

Another major environmental concern is **water consumption**. Water is needed to both cool down many of the electricity-generating sources of the data centres, as well as the many thousand servers within the data centres (Ren, 2023). The initial training of the models is also a water extensive practice, exemplified by the millions of litres of water used to train GPT-3 (Li, P. et al., 2023). GenAI's usage and demand of water is projected to increase significantly in the coming years.

A final environmental risk that should be mentioned here is this technology's reliance on **rare minerals** in the AI supply chain (Dauvergne, 2020). These minerals are often extracted from areas in the Global South, following colonial exploitation patterns, with substantial community and ecological costs (Muldoon and Wu, 2023).

6. Summary of findings from CFA's survey on Danish researchers' use of GenAI

6.1. Introduction to survey

Early 2024, a team of researchers at the Danish Center for Studies in Research and Research Policy (CFA), including two of the authors of the present report (MPS and SPJMH), conducted a nationwide survey among researchers at Danish universities about their use and assessment of GenAI for research purposes. To supplement the findings from the literature study and interviews conducted in relation to this report and described in the previous chapters, we here summarise the main findings of the survey study (see Andersen et al., 2024).[1] First, we briefly describe the survey methodology, before presenting the main findings. Full documentation of the survey, the data gathered, and the results obtained can be found on the survey project's OSF page (OSF, 2024) and in Andersen et al. (2024).

Based on a list of 32 potential use cases of GenAI for research purposes, respondents were asked about whether they had used GenAI for this purpose, whether they know of a close colleague who had used it for this purpose, and how they assessed the use of GenAI for this purpose in terms of research integrity (on a seven-point scale ranging from 'problematic research practice' to 'excellent research practice'). In addition, respondents were asked about their awareness of regulations regarding the use of GenAI in research and were given the opportunity to contextualize their responses in an open text field. The survey was conducted in January and February 2024, i.e. prior to the other work for this report. That means that the use cases included in the survey only partly overlap with the ones described in Chapter 4. Similarly, the definition of GenAI used by the survey is slightly different from the one described in Chapter 2 of this report.

Contact information of all researchers, including PhD students, at the Danish universities was collected via institutional webpages. The survey was subsequently sent to 29,498 individuals, rendering complete responses from 2,527 (8.6%) and partial responses from another 533 (1.8%). For an overview of the study population and sample's demographics, we refer to the survey project's OSF page (OSF, 2024).

6.2. How do researchers based in Denmark use GenAI?

In this chapter we will outline the main results of the survey, related to the respondents' self-reported use of GenAI tools for research purposes. We here focus on their research integrity assessment of the use of GenAI for various purposes. For more details, we refer to the data set and preprint mentioned above.

Figure 6.1. show the respondents' use of GenAI for various research tasks. First and foremost, the usage patterns indicate considerable diversity in the extent to which respondents use GenAI for these tasks. While usage for most use cases is rather limited, it has already become a widespread practice for a few specific use cases, mostly related to language editing (e.g. in proposals, abstracts or research articles).

[1] The figures and parts of the text in this chapter are from Andersen et al., 2024.

Figure 6.1. Share of participants using GenAI for specific use cases. Results are shown by research phase. Blue dots in the right panel show how large a share of respondents that report ever having used AI for the specific use case, while yellow dots show the share of respondents who report that they believe their colleagues use AI for this use case. Horizontal lines in the right panel serve as visual guides only.

For some use cases in the survey, there were hardly any self-reported use of GenAI. This includes use cases such as ‘helping identify potential collaborators’, ‘identifying ethical issues’, and the use of GenAI for peer review purposes. More broadly, for only eight out of the 32 use cases surveyed, at least 25% of respondents indicate to have used GenAI for this purpose. There might be several reasons for the low degree of GenAI use for most research tasks. These reasons might include a lack of awareness that GenAI can be used for this task, lack of skills, perceived ethics/integrity issues, and uncertainty about how others will judge this.

Looking at respondents’ background, the survey showed variation across disciplines and ways of producing knowledge. Despite relative consistency in the use patterns across disciplines, a somewhat larger proportion of respondents from the technical sciences, especially those in the experimental technical sciences, indicate to use GenAI for a higher number of different use cases, some even for more than half of all use cases mentioned in our survey. In contrast, scholars from the humanities indicate to use GenAI tools least frequently. Similarly, the survey found that quantitative social scientists indicate to use GenAI tools for substantially more use cases than their colleagues from the qualitative social sciences. No substantial differences were observed between men and women in their use of GenAI. In terms of academic age, junior scholars use GenAI for more tasks than their senior colleagues.

As might be expected, respondents consistently report that their direct colleagues use GenAI more than they do themselves. For almost all use cases, the difference between reported own use versus reported awareness of colleagues’ use of GenAI is between 30 and 40 per cent. Use cases with notably high relative differences include tasks related to peer review, creation of software code for simulations, and the creation of slide decks for conference talks or other academic events.

In terms of the kind of tools used, most respondents mentioned ChatGPT. A few other general purpose GenAI tools like Google's Bard (now Gemini) were also mentioned, but to a distinctively lesser extent. In addition, several respondents mentioned more specialised tools, usually tailored to specific tasks like Elicit (for literature reviews) and GitHub Copilot (for coding tasks). Interestingly, some respondents also mentioned tools that, in our definition as outlined in Chapter 2 in this report, would not be considered Generative AI, such as grammar or spelling-checker tools.

6.3. What do researchers think about using GenAI – good/bad practice?

While usage of GenAI for most use cases studied in the survey remains relatively limited, respondents’ assessment of these use cases tended to be (mildly) positive. Respondents were asked to assess the use of GenAI for each of the 32 use cases on a 7-point Lickert scale, ranging from ‘problematic research practice’ to ‘excellent research practice’.

Figure 6.2. presents the research integrity scores for each of the use cases surveyed. Assessments varied considerably, with at least some respondents providing assessments on the extreme ends of the spectrum for every use case. However, on average, respondents tended to be fairly positive in their assessments. For most use cases, the average assessment is in the range ‘good’ to ‘excellent’ research practice. This was particularly the case for use cases related to language editing (e.g. in proposal writing, editing of research articles, formatting references) and data analysis (e.g. creating codes for analysis or simulation, pattern recognition, transcription of research recordings). Respondents were also generally positive about use cases related to the translation of research findings to communicate to various audiences (e.g. translating research papers into other languages, creating lay summaries of research findings and creating slide decks for academic events).

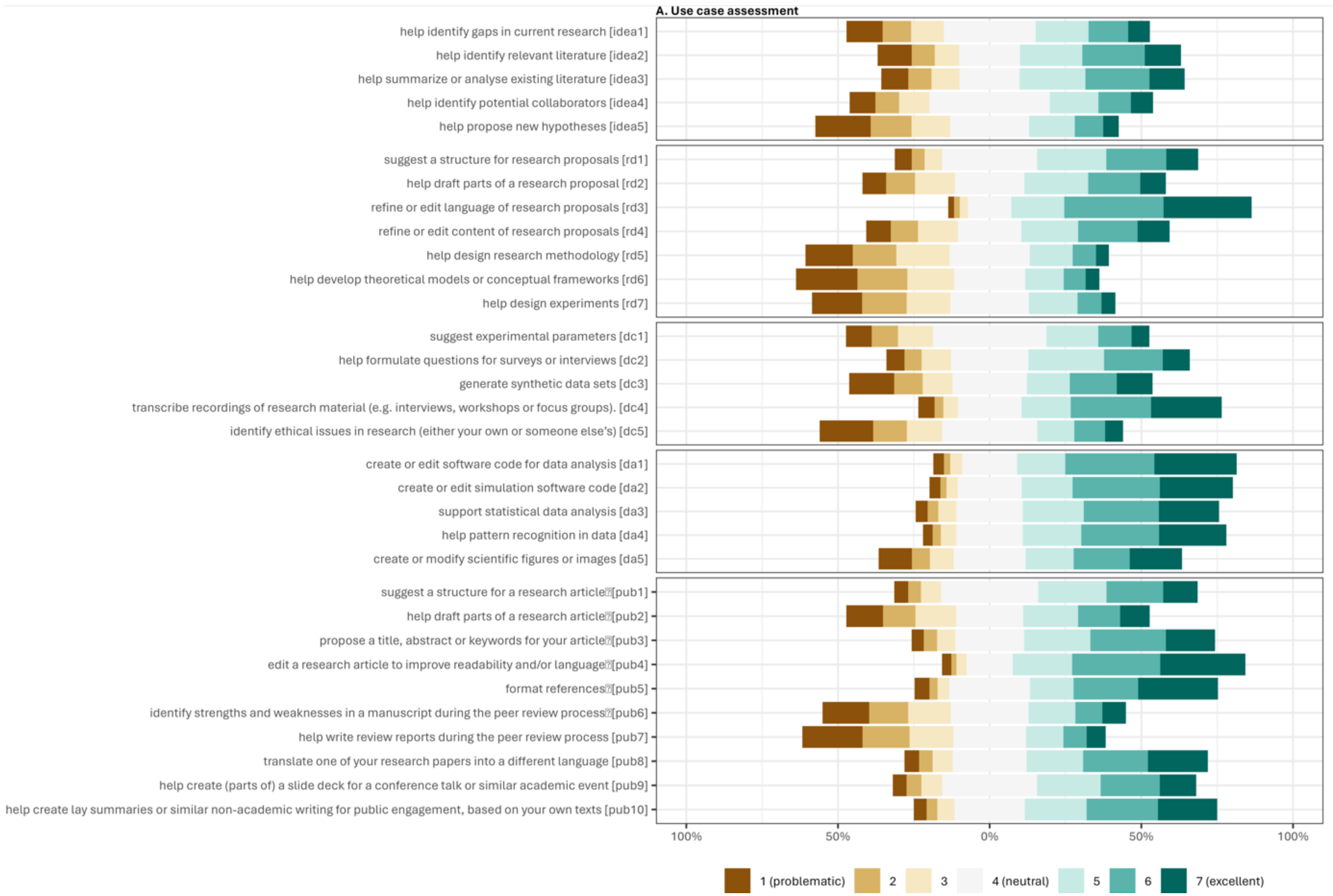


Figure 6.2. Research integrity assessment scores for specific use cases. Results are shown by research phase. Brown bars show the shares of respondents judging the use case as a problematic practice, while green bars show positive assessments.

In contrast, respondents assessed the usage of GenAI for arguably more fundamental tasks related to designing research experiments or theoretical frameworks and critical evaluation of other work during peer review as more problematic. The use cases related to the creation or modification of images and figures, and the creation of synthetic data were particularly contentious. This can likely be explained by the fact that both these cases have different connotations in diverse research fields.

The survey noted that for almost all use cases, the share of respondents indicating a use case to be a good, very good or excellent practice is higher than the share of respondents indicating to have used GenAI for this purpose. This suggests that, while reported use of GenAI is still fairly low, the reason for not using GenAI for more tasks is probably not primarily related to research integrity considerations. This is further underlined by comparing aggregated assessment and usage scores. Respondents reporting the use of GenAI for more use cases, have higher assessment scores and lower variance. Conversely, respondents that had not used GenAI, or only had used it in very few use cases, had much higher disagreement on the assessment of the use cases on average. In the open text fields, many respondents contextualised their research integrity assessments, by indicating that for all use cases, if at all, GenAI should always only be used critically and reflexively, echoing the perspective of the interviewees of this current report.

To identify patterns in the research assessment scores of our respondents, the survey team performed an exploratory factor analysis. This analysis identified three clusters of GenAI perceptions, which were labelled "**GenAI as a work horse**," "**GenAI as a language assistant only",** and "**GenAI as a research accelerator**". Figure 6.3. presents the research integrity assessment and own use of respondents in each of the three clusters.

In the cluster "**GenAI as a work horse**", we find researchers who approve of using GenAI to create and edit software codes for analysis and simulation, to support statistical analysis and to help recognize patterns in data. On the other hand, researchers in this cluster are more sceptical towards using GenAI in the peer review process and in the 'Idea generation phase'. In this cluster, researchers working with qualitative data and researchers from theoretical or conceptual traditions are less well represented.

The second cluster, "**GenAI as a language assistant only**", contains the most sceptical respondents. They generally assess the use of GenAI more negatively than the other clusters, but it is also in this cluster we find the most "neutral" responses (i.e. respondents that might still be undecided about the integrity of using GenAI). While respondents in this cluster give generally more critical assessment scores, they are particularly more sceptical about the use of GenAI for data analysis purposes. The only use cases for which respondents in this cluster tend to be more positive are those related to language editing. There is a strong overrepresentation of scholars from the humanities, theoretical natural scientists, and qualitative social scientists. In comparison, much fewer quantitative social scientists, experimental natural scientists, basic medical scientists, and experimental technical scientists are represented in this cluster.

The third cluster, "**GenAI as a research accelerator**", contains the most outspoken GenAI optimists. Researchers in this cluster have the most positive research integrity assessment for almost all use cases. They seem to put emphasis on the opportunities that GenAI brings, particularly in terms of efficiency gains. In terms of demographics and background, especially junior researchers are well-represented in this cluster.

Even though research integrity assessments of the three clusters are markedly different, the use patterns of respondents in each cluster are highly similar. The use only differs in the degree of usage and not the kind of tasks that GenAI tends to be used for.

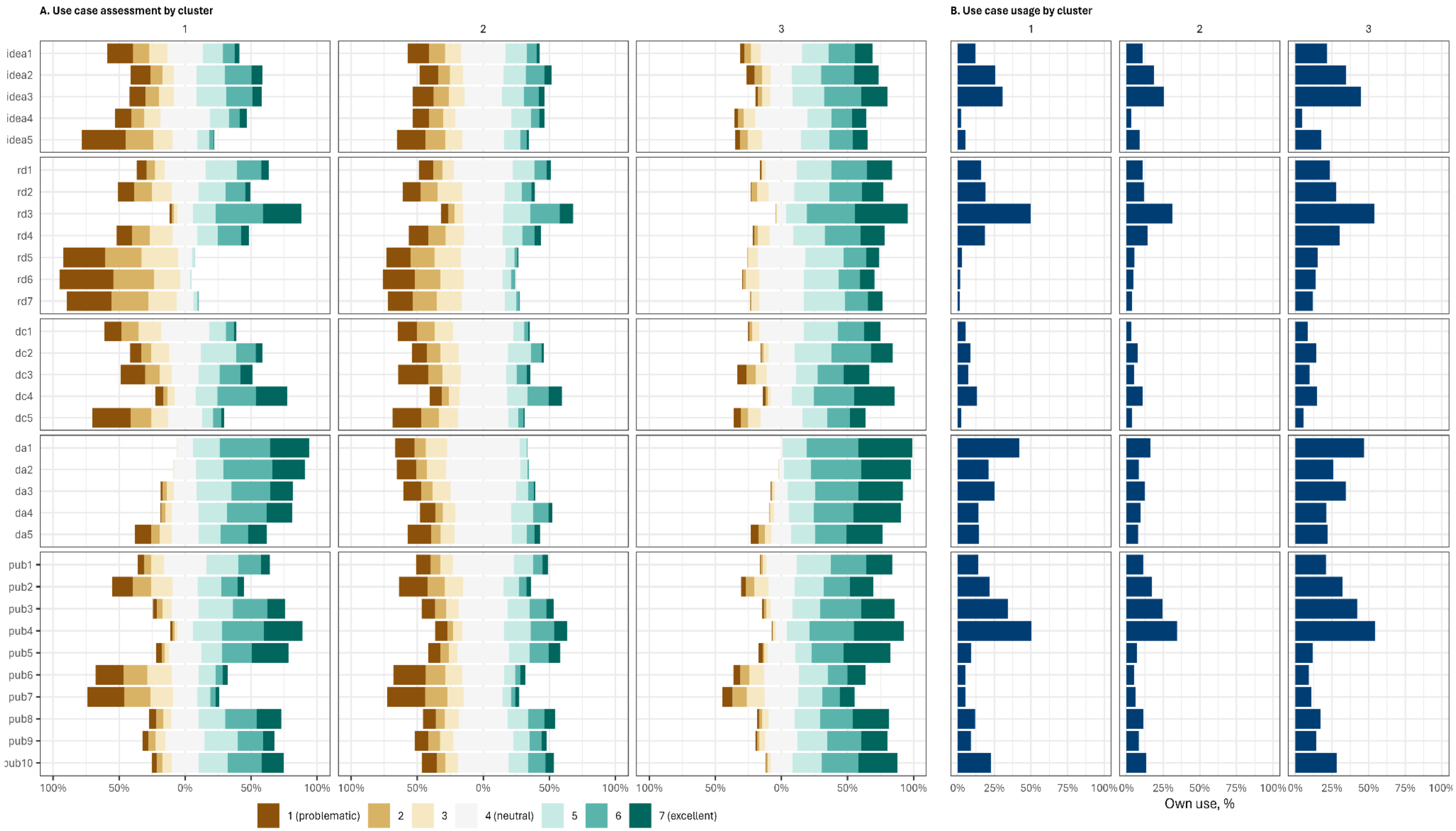


Figure 6.3.: Research integrity assessment responses and own use across all use cases, split by the three clusters identified in the factor analysis.

7. Discussion and conclusion

In this report, we have examined the potential and risks of using GenAI in the different phases of the research process. First, we looked at the evolution of GenAI – from AI, Machine Learning and Deep Learning to different GenAI models like LLMs and GANs. Hereafter, we examined the use of GenAI in the different research phases, using an analytical research process model with five phases: Idea generation and research funding, Research design, Data collection, Data analysis, and Scientific publishing, reporting and dissemination. The analysis was built on both a literature review and an interview study.

The analysis showed that there are substantial differences between research fields when it comes to the use of GenAI models and tools in the research process and similarly substantial variation in the potential of such tools within different fields. Looking at how relevant GenAI is across the five phases of the research process, the study showed that GenAI tools can assist tasks within ‘Idea generation and research funding’ and ‘Scientific publishing, reporting and dissemination’ such as translation, language editing, writing, literature reviews across most fields. In contrast, for ‘Data collection’ and ‘Data analysis’, there are considerable differences between fields in how relevant it is to use GenAI models and tools, due to variations in the studied areas and phenomena as well as differences in epistemic cultures, analytical strategies and the methods employed.

When thinking about the potential of GenAI, it is also important to think about research as one activity among others for researchers. Administration and teaching obligations typically also take up a substantial part of a researcher’s workday. If GenAI could be used to solve more tedious and time-consuming administrative tasks, it would help free up time for research. There likewise seem to be interesting perspectives in integrating GenAI in teaching, something that has been outside the scope of this report to investigate.

When looking at the results presented in this report, especially the concrete tools and use cases, it is important to bear in mind that the report provides a snapshot of the current GenAI landscape, i.e. its state in the summer and early autumn of 2024. The literature that we built on grows rapidly at the moment. As shown above, the literature identified with our search string grew from approximately 600 articles in July 2024 to 750 in the beginning of September 2024. This means that new use cases and new tools as well as new relevant discussions of potential and risks keep being added to this literature, making GenAI use in the research process a tricky, moving target to study.

Although more studies are needed to fully identify the potential of GenAI in the research process within individual fields, we can already now see that there is no one-size-fits-all approach for integrating GenAI in the research process, across research fields and disciplinary contexts. We therefore recommend that concrete strategies for integrating GenAI in different research phases are developed in close interaction with the individual research fields and disciplines, to make these strategies relevant for the research environments and to avoid as many of the pitfalls and risks identified in Chapter 5 as possible. These efforts could include local workshops, focus groups and other mutual learning activities. Similarly, to make policies and guidelines relevant for a particular research community, it is crucial that they are co-created with this community, involving researchers as well as other relevant stakeholders such as learned societies, funders and publishers.

Finally, it is worth noticing that although GenAI can help accelerate research processes in many fields, it is important that the increased speed does not jeopardise the quality of the research produced. As part of a dialogue with the research environments on how to integrate GenAI in the research process, it is therefore crucial to also find ways to co-create policies and guidelines that can help the different communities find the right balance between speed and quantity, on the one side, and thoroughness and quality on the other. As some of our interviewees pointed out, the acceleration of the research process due to GenAI can put other processes, such as the peer review process or academic reward systems, under pressure. These factors should be carefully considered when guiding GenAI's implementation in research.

7.1. Limitations

Like all studies, this one also has its limitations. These include **the scope of the study.** The nine interviews we conducted provided us with invaluable insights, but they were not able to cover all main fields and ways of producing knowledge. More interviews would probably have added more nuances and interesting use cases to the study results.

For the same reasons, we have not been able to fully validate the value of all the use cases listed in this report. A mention of a use case does, therefore, not necessarily mean that it is a very efficient or widely cherished practice within a field, but only that it has been tested and that a peer reviewed article has proposed or reported on the use. As far as possible, we used the literature and the interviews to **validate identified cases**, but to fully validate the usefulness and efficiency of the cases listed in the tables and text in Chapter 4 would require a much stronger engagement process with the research community, for example, through workshops or focus group interviews with representatives from individual disciplines and ways of producing knowledge.

Related to this, it is also important to note that all researchers who were interviewed for this report only **represent themselves**. Although they were also asked about the use of GenAI within their field more broadly, their views do not necessarily represent the general view within their field or discipline. Further, it should also be noted that most of the researchers interviewed were rather **optimistic and positive** towards using GenAI tools in the research process. As shown in Chapter 6, this optimism is not shared by all researchers in Denmark.

Finally, it is also important to mention that there might be a form of **publication bias** in the results of our literature search. Our search strategy was more likely to catch papers that present successful new applications of GenAI, while studies highlighting the risks or challenges related to GenAI – or failed attempts to use GenAI – might be underrepresented in the literature in general and in our search results in particular.

8. References

Agathokleous, E., Rillig, M. C., Peñuelas, J. and Yu, Z. (2024). One hundred important questions facing plant science derived using a large language model. *Trends in Plant Science*, vol. 29, iss. 2, pp. 210-218. https://doi.org/10.1016/j.tplants.2023.06.008

Ali, I., Greifeneder, F., Stamenkovic, J. et al. (2015). Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sens*, 7(12), pp. 16398-16421. https://doi.org/10.3390/rs71215841

Aliaga, D. and Niyogi, D. (2024). Digitizing cities for urban weather: representing realistic cities for weather and climate simulations using computer graphics and artificial intelligence. *Computational Urban Science,* vol. 4, no. 8. https://doi.org/10.1007/s43762-023-00111-z

Altmäe, S., Sola-Leyva, A. and Salumets, A. (2023). Artificial intelligence in scientific writing: a friend or a foe? *Reproductive BioMedicine Online*, vol. 47, iss. 1, pp. 3-9. https://doi.org/10.1016/j.rbmo.2023.04.009

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., et al. (2023). The manifold costs of being a non-native English speaker in science. *PLoS Biology,* 21(7): e3002184. https://doi.org/10.1371/journal.pbio.3002184

Andersen, J. P., Degn, L., Fishberg, R. et al. (2024). Generative Artificial Intelligence (GenAI) in the research process – a survey of researchers' practices and perceptions. http://dx.doi.org/10.31235/osf.io/83whe

Bai, X., Xie, Y., Han, H. and Li, J-R. (2024). Evaluation of Open-Source Large Language Models for Metal-Organic Frameworks Research. *Journal of Chemical Information and Modeling*, vol. 64, iss. 13, pp. 4958-4965. https://doi.org/10.1021/acs.jcim.4c00065

Bano, M., Hoda, R., Zowghi, D. and Treude, C. (2024). Large language models for qualitative research in software engineering: exploring opportunities and challenges. *Automated Software Engineering*, vol. 31, no. 8. https://doi.org/10.1007/s10515-023-00407-8

Bengesi, S., El-Sayed, H., Sarker, K. et al. (2024). Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. *IEEE Access*, vol. 12, pp. 69812-69837. https://doi.org/10.1109/ACCESS.2024.3397775

Bhayana, R. (2024). Chatbots and Large Language Models in Radiology: A Practical Primer for Clinical and Research Applications, *Radiology* 2024; 310(1). https://doi.org/10.1148/radiol.232756

Boiko, D. A., MacKnight, R., Kline, B. and Gomes, G. (2023). Autonomous chemical research with large language models. *Nature*, 624, pp. 570-578. https://doi.org/10.1038/s41586-023-06792-0

Bojic, S., Radovanovic, N., Radovic, M., and Stamenkovic, D. (2024). Could generative artificial intelligence replace fieldwork in pain research? *Scandinavian Journal of Pain*, vol. 24, no. 1. https://doi.org/doi:10.1515/sjpain-2023-0136

Bol, T., de Vaan, M. and van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, vol. 115, no. 19, pp. 4887-4890. https://doi.org/10.1073/pnas.1719557115

Bonnechère, B. (2024). Unlocking the Black Box? A Comprehensive Exploration of Large Language Models in Rehabilitation. *American Journal of Physical Medicine & Rehabilitation*, 103(6), pp. 532-537. http://dx.doi.org/10.1097/PHM.0000000000002440

Cao, Y., Li, S., Liu, Y. et al. (2023). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. *ArXiv.* https://doi.org/10.48550/arXiv.2303.04226.

Cao, X., Xu, W., Zhao, J. et al. (2024). Research on Large Language Model for Coal Mine Equipment Maintenance Based on Multi-Source Text. *Applied Sciences,* vol. 14, iss. 7, 2946. https://doi.org/10.3390/app14072946

Cappelli, O., Aliberti, M., and Praino, R. (2024). The 'Implicit Intelligence' of artificial intelligence. Investigating the potential of large language models in social science research. *Political Research Exchange*, vol. 6, iss. 1. https://doi.org/10.1080/2474736X.2024.2351794

Chatterjee, S., Bhattacharya, M., Pal, S. et al. (2023). ChatGPT and large language models in orthopedics: from education and surgery to research. *Journal of Experimental Orthopaedics*, vol. 10, iss. 1, 128. https://doi.org/10.1186/s40634-023-00700-1

Chen, Y., Zhang, H., Xu, C. et al. (2024). Research on seawater dissolved oxygen prediction model based on improved generative adversarial networks. *Ocean Modelling*, vol. 191, 102404. https://doi.org/10.1016/j.ocemod.2024.102404

Chui, K. T., Liu, R. W., Zhao, M., and Pablos, P. O. D. (2020). Predicting Students' Performance With School and Family Tutoring Using Generative Adversarial Network-Based Deep Support Vector Machine. *IEEE Access*, 8, pp. 86745-86752. https://doi.org/10.1109/ACCESS.2020.2992869

Cole, V. and Boutet, M. (2023). ResearchRabbit. T*he Journal of the Canadian Health Libraries Association*, 44(2), pp. 43-47. https://doi.org/10.29173%2Fjchla29699

Cornell University Taskforce (2023). Generative AI in Academic Research: Perspectives & Cultural Norms. https://it.cornell.edu/sites/default/files/itc-drupal10-files/Generative%20AI%20in%20Research_%20Cornell%20Task%20Force%20Report-Dec2023.pdf

Coskun, A. (2024). AI supercharges data center energy use – straining the grid and slowing sustainability efforts. *The Conversation*. 15th of July 2024. https://theconversation.com/ai-supercharges-data-center-energy-use-straining-the-grid-and-slowing-sustainability-efforts-232697 (Accessed 18/09-24)

Courtenay, L. A. and González-Aguilera, D. (2020). Geometric Morphometric Data Augmentation Using Generative Computational Learning Algorithms. *Applied Sciences*, 10(24), 9133. https://doi.org/10.3390/app10249133

Cui, W., Xiao, M., Wang, L. et al. (2024). Automated taxonomy alignment via large language models: bridging the gap between knowledge domains. *Scientometrics.* https://doi.org/10.1007/s11192-024-05111-2

Dashkevych, O. and Portnov, B. A. (2024). How can generative AI help in different parts of research? An experiment study on smart cities' definitions and characteristics. *Technology in Society*, vol. 77. https://doi.org/10.1016/j.techsoc.2024.102555

Dauvergne, P. (2020). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Review of International Political Economy*, vol. 29, iss. 3, pp. 696–718. https://doi.org/10.1080/09692290.2020.1814381

DeepMind (2024): https://deepmind.google/discover/blog/alphaproteo-generates-novel-proteins-for-biology-and-health-research/ (Accessed 23/09-24)

Dengel, A., Gehrlein, R., Fernes, D. et al. (2023). Qualitative Research Methods for Large Language Models: Conducting Semi-Structured Interviews with ChatGPT and BARD on Computer Science Education. *Informatics*, vol. 10, iss. 4. https://doi.org/10.3390/informatics10040078

Erler, A. (2023). Publish with AUTOGEN or Perish? Some Pitfalls to Avoid in the Pursuit of Academic Enhancement via Personalized Large Language Models. *The American Journal of Bioethics*, 23(10), pp. 94–96. https://doi.org/10.1080/15265161.2023.2250291

Esteva, A., Kuprel, B., Novoa, R. et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, pp. 115–118. https://doi.org/10.1038/nature21056

Formanek, M. (2024). Exploring the potential of large language models and generative artificial intelligence (GPT): Applications in Library and Information Science. *Journal of Librarianship and Information Science*, 0(0). https://doi.org/10.1177/09610006241241066

García-Peñalvo, F. and Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. *International Journal of Interactive Multimedia and Artificial Intelligence.* https://doi.org/10.9781/ijimai.2023.07.006

Giuffrè, M. and Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *npj Digital Medicine*, 6(1), 186. https://doi.org/10.1038/s41746-023-00927-3

Godwin, R., DeBerry, J. J., Wagener, B. M. et al. (2024). Grant drafting support with guided generative AI software. *SoftwareX*, vol. 27, 101784. https://doi.org/10.1016/j.softx.2024.101784

Gray, A. (2024). ChatGPT “contamination": estimating the prevalence of LLMs in the scholarly literature. *arXiv preprint arXiv.* https://doi.org/10.48550/arXiv.2403.16887

Grimaldi, G. and Ehrler, B. (2023). AI et al.: Machines Are About to Change Scientific Publishing Forever. *ACS Energy Letters*, vol. 8, iss. 1, pp. 878-880. https://doi.org/10.1021/acsenergylett.2c02828

Guo, D., Yue, A., Ning, F. et al. (2023). A Study Case of Automatic Archival Research and Compilation using Large Language Models. *IEEE International Conference on Knowledge Graph*, pp. 52-59. https://doi.org/10.1109/ICKG59574.2023.00012

Guo, M., Wu, F., Jiang, J., et al., (2023). Investigations on Scientific Literature Meta Information Extraction Using Large Language Models. *2023 IEEE International Conference on Knowledge Graph (ICKG),* (pp. 249-254). IEEE. https://doi.ieeecomputersociety.org/10.1109/ICKG59574.2023.00036

Hämäläinen, P., Tavast, M., and Kunnari, A. (2023). Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1-19. https://doi.org/10.1145/3544548.3580688

Hicks, M.T., Humphries, J. and Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology,* vol. 26, no. 38. https://link.springer.com/article/10.1007/s10676-024-09775-5

Hirosawa, T. and Shimizu, T., (2024). Enhancing English Presentation Skills with Generative Artificial Intelligence: A Guide for Non-native Researchers. *Medical Science Educator.* https://doi.org/10.1007/s40670-024-02078-w

Hosseini, M. and Horbach, S. P. J. M. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review*, 8(1), 4. https://doi.org/10.1186/s41073-023-00133-5

Hosseini, M., Horbach, S. P. J. M., Holmes, K. L. and Ross-Hellauer, T. (2024). Open Science at the Generative AI Turn: An Exploratory Analysis of Challenges and Opportunities. *SocArXiv*. https://doi.org/10.31235/osf.io/zns7g

Hu, Y., Mai, G., Cundy, C. et al. (2023). Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science*, vol. 37, iss. 11, pp. 2289-2318. https://doi.org/10.1080/13658816.2023.2266495

Hwang, S., Lim, J. S., Lee, R. W. et al. (2023). Is ChatGPT a "Fire of Prometheus" for Non-Native English-Speaking Researchers in Academic Writing? *Korean Journal of Radiology*, 24(10), pp. 952-959. https://doi.org/10.3348/kjr.2023.0773

International Energy Agency (2024). "Electricity 2024. Analysis and forecast to 2026". https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf (Accessed 18/09-24)

Islam, S., Javeed, D., Saeed, M. S. et al. (2024). Generative AI and Cognitive Computing-Driven Intrusion Detection System in Industrial CPS. *Cognitive Computation*, vol. 16, pp. 2611–2625. https://doi.org/10.1007/s12559-024-10309-w

Jeong, C-H. and Yi, M. Y. (2023). Correcting rainfall forecasts of a numerical weather prediction model using generative adversarial networks. *The Journal of Supercomputing*, vol. 79, 1289–1317. https://doi.org/10.1007/s11227-022-04686-y

Jiang, S., Evans-Yamamoto, D., Bersenev, D. et al. (2024). ProtoCode: Leveraging large language models (LLMs) for automated generation of machine-readable PCR protocols from scientific publications. *Society for Laboratory Automation and Screening Technology*, vol. 29, iss. 3. https://doi.org/10.1016/j.slast.2024.100134

Karagiorgi, G., Kasieczka, G., Kravitz, S. et al. (2022). Machine learning in the search for new fundamental physics. *Nature Review Physics*, 4, pp. 399–412. https://doi.org/10.1038/s42254-022-00455-1

Khoshkebari, M. P. (2023). How 'Attention is All You Need' Changed The Course Of Generative AI. https://medium.com/@pasha.khoshkeba/how-attention-is-all-you-need-changed-the-course-of-generative-ai-25b8c24ca430 (Accessed 26/08-24)

Kim, J., Suh, S., Chilton, L. B. and Xia, H. (2023). Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. *Association for Computing Machinery*, pp. 115-135. https://doi.org/10.1145/3563657.3595996

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications, Limited. http://ebookcentral.proquest.com/lib/asb/detail.action?docID=1712661

Kitchin, R. (2021). *Data lives: How data are made and shape our world*. Bristol University Press.

Klinger, J., Mateos-Garcia, J. and Stathoulopoulos, K. (2022). A narrowing of AI research? *arXiv preprint arXiv.* https://doi.org/10.48550/arXiv.2009.10385

Kobak, D., Márquez, R. G., Horvát, E. Á. and Lause, J. (2024). Delving into ChatGPT usage in academic writing through excess vocabulary. *arXiv preprint arXiv*. https://doi.org/10.48550/arXiv.2406.07016

Latona, G. R., Ribeiro, M. H., Davidson, T. R. et al. (2024). The AI Review Lottery: Widespread AI-Assisted Peer Reviews Boost Paper Scores and Acceptance Rates. *arXiv preprint arXiv*. https://doi.org/10.48550/arXiv.2405.02150

Laudel, G. (2024). Where do field-specific notions of research quality come from? *Research Evaluation*, vol. 33, iss. 1, rvae027. https://doi.org/10.1093/reseval/rvae027

Li, P., Yang, J., Islam, M. A., and Ren, S. (2023). Making ai less" thirsty": Uncovering and addressing the secret water footprint of ai models. *arXiv preprint*, https://doi.org/10.48550/arXiv.2304.03271

Li, X., Huang, T., Cheng, K., Qiu, Z. and Sichao, T. (2022). Research on anomaly detection method of nuclear power plant operation state based on unsupervised deep generative model. *Annals of Nuclear Energy*, vol. 167, 108785. https://doi.org/10.1016/j.anucene.2021.108785

Linegar, M., Kocielnik, R. and Alvarez, R. M. (2023). Large language models and political science. *Frontiers in Political Science*, vol. 5. https://doi.org/10.3389/fpos.2023.1257092

Lissak, S., Ophir, Y., Tikochinski, R. et al. (2024). Bored to death: Artificial Intelligence research reveals the role of boredom in suicide behavior. *Frontiers in Psychiatry*, vol. 15. https://doi.org/10.3389/fpsyt.2024.1328122

Liu, J., Xia, X., Peng, X. et al. (2022). Research on ECG Signal Classification Based on Data Enhancement of Generative Adversarial Network. *Artificial Intelligence and Security*, vol. 13338, Springer, pp. 405-419. https://doi.org/10.1007/978-3-031-06794-5_33

Lozano, A., Fleming, S. L., Chiang, C-C. and Shah, N. (2023). Clinfo.ai: An Open-Source Retrieval-Augmented Large Language Model System for Answering Medical Questions using Scientific Literature. *Biocomputing*, pp. 8-23. https://doi.org/10.1142/9789811286421_0002

MacNeil, S., Tran, A., Hellas, A. et al. (2023). Experiences from Using Code Explanations Generated by Large Language Models in a Web Software Development E-Book. *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, pp. 931-937. https://doi.org/10.1145/3545945.3569785

Mann, S. P., Earp, B. D., Møller, N. et al. (2023). AUTOGEN: A Personalized Large Language Model for Academic Enhancement—Ethics and Proof of Principle. *The American Journal of Bioethics*, 23(10), 28–41. https://doi.org/10.1080/15265161.2023.2233356

Moon, J., Lee, U., Koh, J. et al. (2024). Generative Artificial Intelligence in Educational Game Design: Nuanced Challenges, Design Implications, and Future Research. *Technology, Knowledge and Learning*. http://dx.doi.org/10.1007/s10758-024-09756-z

Moussad, B., Roche, R. and Bhattacharya, D. (2023). The transformative power of transformers in protein structure prediction. *Proceedings of the National Academy of Sciences*, vol. 120, no. 32. e2303499120. https://doi.org/10.1073/pnas.2303499120

Muhammad, A., Salman, Z., Lee, K. and Han, D. (2023). Harnessing the power of diffusion models for plant disease image augmentation. *Frontiers in Plant Science*, vol. 14. https://doi.org/10.3389/fpls.2023.1280496

Muldoon, J. and Wu, B. A. (2023) Artificial Intelligence in the Colonial Matrix of Power. *Philosophy and Technology*, vol. 36, no 80. https://doi.org/10.1007/s13347-023-00687-8

OSF (Open Science Framework) (2024). https://osf.io/6htfs/ (Accessed 24/09-24)

Pang, S. M., Cao, J. X., Jian, M. Y. et al. (2022). BR-GAN: A Pedestrian Trajectory Prediction Model Combined With Behavior Recognition. *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, iss. 12, pp. 24609-24620. https://doi.org/10.1109/TITS.2022.3193442

Park, Y. J., Kaplan, D., Ren, Z. et al. (2024). Can ChatGPT be used to generate scientific hypotheses? *Journal of Materiomics*, vol. 10, iss. 3, pp. 578-584. https://doi.org/10.1016/j.jmat.2023.08.007

Peng, C., Zhao, X. and Xia, G. (2023): Research on Colorization of Qinghai Farmer Painting Image Based on Generative Adversarial Networks. *Association for Computing Machinery*, pp. 495-503. https://doi.org/10.1145/3590003.3590094

Pividori, M. and Greene, C. S. (2024). A publishing infrastructure for Artificial Intelligence (AI)-assisted academic authoring. *Journal of the American Medical Informatics Association*, Volume 31, Issue 9, September 2024, pp. 2103–2113. https://doi.org/10.1093/jamia/ocae139

Potter, W. (2024). An academic publisher has struck an AI data deal with Microsoft – without their authors' knowledge. *The Conversation*. 23rd of July 2024. https://theconversation.com/an-academic-publisher-has-struck-an-ai-data-deal-with-microsoft-without-their-authors-knowledge-235203

Pride, D., Cancellieri, M. and Knoth, P. (2023). CORE-GPT: Combining Open Access Research and Large Language Models for Credible, Trustworthy Question Answering. *Linking Theory and Practice of Digital Libraries,* TPDL 2023, Lecture Notes in Computer Science, vol. 14241. Springer. http://dx.doi.org/10.1007/978-3-031-43849-3_13

Raja, H., Mumawar, A., Mylonas, N. et al. (2024). Automated Category and Trend Analysis of Scientific Articles on Ophthalmology Using Large Language Models: Development and Usability Study. *Journal of Medical Internet Research*, vol. 8. https://doi.org/10.2196/52462

Ren, S. (2023): How much water does AI consume? The public deserves to know. https://oecd.ai/en/wonk/how-much-water-does-ai-consume (Accessed 19/09-24)

Rengers, T. A., Thiels, C. A. and Salehinejad, H. (2024). Academic Surgery in the Era of Large Language Models: A Review. *JAMA Surgery*. 159(4), pp. 445-450. https://doi.org/10.1001/jamasurg.2023.6496

Resnik, D. B. and Hosseini, M. (2023). The Impact of AUTOGEN and Similar Fine-Tuned Large Language Models on the Integrity of Scholarly Writing. *The American Journal of Bioethics*, 23(10), pp. 50–52. https://doi.org/10.1080/15265161.2023.2250276

Robson, J. F., Denholm, S. J. and Coffey, M. (2021). Automated Processing and Phenotype Extraction of Ovine Medical Images Using a Combined Generative Adversarial Network and Computer Vision Pipeline. *Sensors*, vol. 21, iss. 21, 7268. https://doi.org/10.3390/s21217268

Romanelli, V., Cerchia, C. and Lavecchia, A. (2024). Unlocking the Potential of Generative Artificial Intelligence in Drug Discovery. *Applications of Generative AI,* pp. 37-63. https://doi.org/10.1007/978-3-031-46238-2_3

Rowe, N. (2023). Millions of Workers Are Training AI Models for Pennies, *Wired*, Oct 16. https://www.wired.com/story/millions-of-workers-are-training-ai-models-for-pennies/ (Accessed 19/09-24)

Rui, J.and Qiang, N. (2022). Research on textile defects detection based on improved generative adversarial network. *Journal of Engineered Fibers and Fabrics,* vol. 17. https://doi.org/10.1177/15589250221101382

Schmidt, P., Arlt, S., Ruiz-Gonzales, C. et al. (2024). Virtual reality for understanding artificial-intelligence-driven scientific discovery with an application in quantum optics. *Machine Learning: Science and Technology*, vol. 5, no. 3. https://iopscience.iop.org/article/10.1088/2632-2153/ad5fdb

Sharma, A. (2024). https://medium.com/@akshitsharma105/celebrating-generative-ai-unveiling-its-core-and-primary-applications-a7cc59efd10a (Accessed 20/09-24)

Shriram, J. and Sreekala, S. P. K. (2023). ZINify: Transforming Research Papers into Engaging Zines with Large Language Models. *Association for Computing Machinery*, article 117, pp. 1-3. https://doi.org/10.1145/3586182.3625118

Shyr, C., Grout, R. W., Kennedy, N. et al. (2024). Leveraging artificial intelligence to summarize abstracts in lay language for increasing research accessibility and transparency. *Journal of the American Medical Informatics Association,* vol. 31, iss. 10, pp. 2294-2302. https://doi.org/10.1093/jamia/ocae186

Sindhura, D. N., Pai, R. M., Bhat, S. N., and Pai, M. M. M. (2024). Deep learning-based automated spine fracture type identification with Clinically validated GAN generated CT images. *Cogent Engineering*, vol. 11, iss. 1. https://doi.org/10.1080/23311916.2023.2295645

Stefano, F. (2023): *AI Evolution: From Basics*. https://blog.stefanofilippone.com/ai-evolution-from-basics-3007caa76607 (Accessed 20/09-24)

Stock, K., Jones, C. B., Russell, S. et al. (2022). Detecting geospatial location descriptions in natural language text. *International Journal of Geographical Information Science*, vol. 36, iss. 3, pp. 547-584. https://doi.org/10.1080/13658816.2021.1987441

Tang, Y., Moretti, R. and Meiler, J. (2024). Recent Advances in Automated Structure-Based De Novo Drug Design. *Journal of Chemical Information and Modeling*, vol. 64, iss. 6. https://doi.org/10.1021/acs.jcim.4c00247

The Royal Society (2024). Science in the age of AI. How artificial intelligence is changing the nature and method of scientific research. https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai-report.pdf

Thelwall, M. (2024). Can ChatGPT evaluate research quality? *Journal of Data and Information Science*, vol. 9. https://doi.org/10.2478/jdis-2024-0013

Urooj, U., Al-Rimy, B. A. S., Zainal, A. B. et al. (2022). Addressing Behavioral Drift in Ransomware Early Detection Through Weighted Generative Adversarial Networks. *IEEE Access*, vol. 12, pp. 3910-3925. https://doi.org/10.1109/ACCESS.2023.3348451

Vaswani, A., Shazeer, N., Parmar, N. et al. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.1706.03762

Wallace, B., Nymoen, K., Torresen, J., and Martin, C. P. (2024). Breaking from realism: exploring the potential of glitch in AI-generated dance. *Digital Creativity*, *35*(2), pp. 125–142. https://doi.org/10.1080/14626268.2024.2327006

Walton, R. O., and Watkins, D. V. (2024). The use of generative AI in research: a production management case study from the aviation industry. *Journal of Marketing Analytics*. https://doi.org/10.1057/s41270-024-00317-y

Wang, H., Tao, G., Ma, J. et al. (2022). Predicting the Epidemics Trend of COVID-19 Using Epidemiological-Based Generative Adversarial Networks. *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, iss. 2, pp. 276-288. https://doi.org/10.1109/JSTSP.2022.3152375

Wang, S., Hu, T., Xiao, H. et al. (2024). GPT, large language models (LLMs) and generative artificial intelligence (GAI) models in geospatial science: a systematic review. *International Journal of Digital Earth*, vol. 17, iss. 1. https://doi.org/10.1080/17538947.2024.2353122

Watson, J.L., Juergens, D., Bennett, N.R. et al. (2023). De novo design of protein structure and function with RFdiffusion. *Nature,* 620, pp. 1089–1100. https://doi.org/10.1038/s41586-023-06415-8

Whitfield, S. and Hofmann, M. A. (2023). Elicit: AI literature review research assistant. *Public Services Quarterly*, vol. 19, iss. 3, pp. 201–207. https://doi.org/10.1080/15228959.2023.2224125

Wollin-Giering, S., Hoffmann, M., Höfting, J., and Ventzke, C. (2024). Automatic Transcription of English and German Qualitative Interviews. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, vol. 25, no. 1. https://doi.org/10.17169/fqs-25.1.4129

Wu, H., He, W., Li, X. and Liang, Y. (2023). Research on Ethnic Pattern Generation Based on Generative Adversarial Networks. *15th International Conference on Advanced Computational Intelligence (ICACI)*, pp.1-6. https://doi.org/10.1109/ICACI58115.2023.10146174

Yu, J., Xu, X., Gao, F. et al. (2021). Toward Realistic Face Photo–Sketch Synthesis via Composition-Aided GANs. *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4350-4362. https://doi.org/10.1109/TCYB.2020.2972944

Zenni, R. D. and Andrew, Nigel R. (2023). Artificial Intelligence text generators for overcoming language barriers in ecological research communication. *Austral Ecology*, 48, pp. 1225-1229. https://doi.org/10.1111/aec.13417

Zhang, Y., Sivarathri, S. S. and Calyam, P. (2020). ScholarFinder: Knowledge Embedding Based Recommendations using a Deep Generative Model. *IEEE Sixth International Conference on Big Data Computing Service and Applications*, pp. 88-95. https://doi.org/10.1109/BigDataService49289.2020.00021

Zhao, S., Chen, S., Zhou, J. et al. (2024). Potential to transform words to watts with large language models in battery research. *Cell Reports Physical Science*, vol. 5, iss. 3. https://doi.org/10.1016/j.xcrp.2024.101844

Zheng, X., Li, J., Lu, M. and Wang, F-Y. (2024). New Paradigm for Economic and Financial Research With Generative AI: Impact and Perspective. *IEEE Transactions on Computational Social Systems,* vol. 11, no. 3, pp. 3457-3467. https://doi.org/10.1109/TCSS.2023.3334306

9. Appendixes

9.1. The full search string for the literature search in Web of Science

Database:	Web of Science Core Collection
Search dates:	last five years (2019 – search date)
Search domain:	title, author key words, key words plus
Search date:	Sept 4, 2024.

Search string:
TI= (("Generative Adversarial" OR "generative ai" OR "genai" OR "generative artificial intelligence" OR "generative model*" OR LLM* OR "Large Language Model*" OR "Large Vision Model*") AND (scien* OR research* OR scholar* OR academ*)) OR AK=(("Generative Adversarial" OR "generative ai" OR genai OR "generative artificial intelligence" OR "generative model*" OR LLM* OR "Large Language Model*" OR "Large vision model*") AND (scien* OR research* OR scholar* OR academ*)) OR KP=(("Generative Adversarial" OR "generative ai" OR genai OR "generative artificial intelligence" OR "generative model*" OR LLM* OR "Large Language Model*" OR "Large vision model*") AND (scien* OR research* OR scholar* OR academ*))

Number of results:	760
Number of results in English:	753

9.2. Interviewees – short bios

Adam Hulman is an Associate Professor at the Department of Public Health at Aarhus University and a Senior Data Scientist at Steno Diabetes Center Aarhus. He holds a PhD in diabetes epidemiology from the University of Szeged. Before entering his current position, he was a postdoctoral researcher funded by the Danish Diabetes Academy. Hulman is the principal investigator of the Machine Learning & Clinical Prediction Lab, an interdisciplinary research group at Steno Diabetes Center Aarhus.

Daniel Russo is an Associate Professor of Software Engineering at the Department of Computer Science at Aalborg University in Copenhagen. Prior to this, he was a postdoctoral researcher at Lero and University College Cork. He holds a PhD in Computer Science and Engineering from the University of Bologna. Daniel Russo has conducted research on topics such as empirical software engineering, AI adoption within the software development lifecycle and advanced research methodologies in software engineering. He is a member at the Danish Young Academy of Technology, Science, and Innovation. He also organizes the Annual Copenhagen Symposium on Human-Centered Software Engineering AI and is likewise affiliated with the Pioneer Centre for Artificial Intelligence.

David Stott is an archaeologist and data scientist at Moesgaard Museum in Aarhus. He has a background in archaeology, geospatial computing, and a PhD in Computer Science. In his archaeological research, he works with large remote sensing datasets, such as Lidar, satellite, multispectral and hyperspectral imagery. Additionally, David Stott does research on the use of artificial intelligence and machine learning in archaeology. He is currently leading a project at Moesgaard Museum that utilizes non-generative AI to map burial mounds in the Danish landscape with the help of old maps of Denmark and citizen science.

Jes Frellsen is an Associate Professor of Machine Learning and Signal Processing at the Technical University of Denmark. He has worked as a postdoctoral researcher both at the University of Cambridge, where he was part of the Machine Learning Group in the Engineering Department, and at the Bioinformatics Centre at Copenhagen University. He holds a PhD in Bioinformatics from the University of Copenhagen. Jes Frellsen works on statistical machine learning with a particular interest in, among others, Generative AI, Deep Learning and Deep Generative Models, and how they can be applied in bioinformatics. He is also affiliated with the Pioneer Centre for Artificial Intelligence.

Johannes Bjerva is a Professor at the Department of Computer Science at Aalborg University. He holds a PhD in Natural Language Processing (NLP) from the University of Groningen. Johannes Bjerva's research interests include the interaction of NLP and linguistics, NLP for low-resource languages, as well as LLMs and their potential in the education and healthcare sectors. He is a fellow of The Young Academy under the Royal Danish Academy of Sciences and Letters. He is also affiliated with the Pioneer Centre for Artificial Intelligence.

Mads Rosendahl Thomsen is a Professor of Comparative Literature at Aarhus University. He also holds a PhD in Comparative Literature from Aarhus University. Mads Rosendahl Thomsen's research focusses on world literature, and he is chair of the Book Panel of the Danish Ministry of Culture. He is also Director of the Center for Language Generation and AI at Aarhus University, and editor of the journal Orbis Litterarum. He is currently co-leading a research project on computational research of literacy fiction and narratives.

Simon Rasmussen is a Professor at Novo Nordisk Foundation Center for Basic Metabolic Research at the University of Copenhagen. He holds a PhD in Systems Biology from the Technical University of Denmark. Simon Rasmussen has been involved in research that transforms large and complex datasets from biological experiments into knowledge and biological understanding. He is currently leading research projects on machine and deep learning for analysis and integration of multi-omics and multi-modal data within cardiometabolic disease. He is also affiliated with the Novo Nordisk Foundation Center for Genomic Mechanism of Disease at Harvard University and the Pioneer Centre for Artificial Intelligence in Denmark.

Timothy Jenkins is Head of Data Science and an Associate Professor at the Technical University of Denmark's (DTU) Department of Bioengineering. He holds a PhD in Biology and Biomedical Sciences from the University of Cambridge. His research focuses on using machine learning to discover and understand protein-based binders, such as antibodies. Timothy Jenkins is leading the Digital Biotechnology Lab at the Technical University of Denmark, a research group which focuses on leveraging the power of data science, AI, and automation to discover and understand protein-protein interactions. He is also a visiting Professor at Novo Nordisk Foundation Center for Biosustainability and is affiliated with the Pioneer Centre for Artificial Intelligence.

9.3. Interview guide

Interview Guide, Expert Interviews, DFiR project on "Using Generative AI across different research phases – potentials and risks" (*last modified: 12 July 2024)*

0. **Preamble**

- Welcome *(2 mins)*
  - Thank you for taking the time … + short intro to interviewer(s) (MPS and xx)
- Recording and consent *(5 mins)*
  - Have they given their consent?
  - Start recording
- Introduction to the project *(5 mins)*
  - What we are doing – who we are, the project and for whom (DFiR: Danmarks Forsknings- og Innovationspolitiske Råd / The Danish Council for Research and Innovation Policy, Ministry of Higher Education and Science)
  - Short intro to today's interview & what will happen after the interview

1. **Their research, expertise** *(5 mins)*
   a. Could you start by telling us a bit about your own research; where you work, your discipline, what you do research on?

2. **Personal use (cases) of Generative AI** *(5-10 mins)*
   a. Could you tell me/us about your experience with generative AI in research?
   b. Do you use GenAI in your own research?
   / What do you work with, and for what tasks/purposes?
   c. What is GenAI within your field?
   / (for more technical experts: GenAI vs. Machine Learning / Deep Learning or AI in general)

3. **Use (cases) of Generative AI in your field?** *(5-10 mins)*
   a. How is Generative AI used within your field
   / The same way as you use it or in other ways?
   What are other applications of generative AI in your field? What do you think about those?
   b. How do you see the development of GenAI over the last years within your field? (development of AI technologies/tools and/or developments in use/application)

4. **GenAI in different research phases** *(10 mins)*
   If we divide the research process into five phases (NB research phases and tasks vary between disciplines/fields): 1) Idea generation and research application work, 2) Research design, 3) Data collection, 4) Data analysis, 5) Scientific publishing, reporting, and dissemination.
   a. In which of these phases are GenAI most used now within your field, in your experience?
   b. Where do you think it could/should be used more? Where (in which phases and for which tasks) do you see the biggest potential for using GenAI within your field?
   / Are there any tasks/phases in your research, or research in your field, that this model is missing? Is there potential in the use of generative AI there?
   c. Are there any risks connected to using GenAI for these tasks? Do you have any other concerns about using GenAI for these tasks?
   / (risks and concerns for whom/what) Whom/what would these risks affect?

5. **Planning for the future** *(5 mins)*

a. How can we better use GenAI – now and in the future? How can we maximize the potential and minimize the risks?
/ Do we need guidelines, training, policies?
b. We have seen a rapid development of GenAI in the last years, how do you see things changing in the future within your field of research? How do we best deal with these ongoing changes and their affordances (and then changes in risks and benefits)?

6. **Policies, guidelines, and training** *(5 mins)*
   a. Are there any regulations of the use of GenAI within your field? Should there be rules/guidelines/etc. on what GenAI can/cannot be used for?

   b. How do you organize training within the use of GenAI for research purposes within your field? Formal/non-formal training? How should training ideally be organized (also considering the accelerated pace of the development of GenAI?

7. **Additional things** *(3-5 mins)*
   a. Before we end, is there anything you would like to add to what we have talked about – anything more that you think is important for us and DFiR to think about when talking about "Generative AI across different research phases – potentials and risks"

8. **Conclusion** *(3-5 mins)*
   a. Thank you for the interview …
   b. What will happen next – and when they can expect to hear from us