Perspective

Author for correspondence:
Denis Newman-Griffis
e-mail: d.r.newman-griffis@sheffield.ac.uk

AI Thinking: a framework for rethinking artificial intelligence in practice

Denis Newman-Griffis[1,2,3]

[1]Centre for Machine Intelligence, The University of Sheffield, Sheffield S1 3JD, UK
[2]Information School, The University of Sheffield, Sheffield S10 2AH, UK
[3]Research on Research Institute, London WC1E 6JA, UK

DN-G, 0000-0002-0473-4226

Artificial intelligence is transforming the way we work with information across disciplines and practical contexts. A growing range of disciplines are now involved in studying, developing and assessing the use of AI in practice, but these disciplines often employ conflicting understandings of what AI is and what is involved in its use. New, interdisciplinary approaches are needed to bridge competing conceptualizations of AI in practice and help shape the future of AI use. I propose a novel conceptual framework called AI Thinking, which models key decisions and considerations involved in AI use across disciplinary perspectives. AI Thinking addresses five practice-based competencies involved in applying AI in context: motivating AI use, formulating AI methods, assessing available tools and technologies, selecting appropriate data and situating AI in the sociotechnical contexts it is used in. A hypothetical case study is provided to illustrate the application of AI Thinking in practice. This article situates AI Thinking in broader cross-disciplinary discourses of AI, including its connections to ongoing discussions around AI literacy and AI-driven innovation. AI Thinking can help to bridge between the work of diverse disciplines, contexts and actors in the AI space, and shape AI efforts in education, industrial development and policy.

1. Introduction

Artificial intelligence has been positioned as a key driver of global change through a fourth industrial revolution, and AI advances are actively transforming how we process, analyse and learn from information [1–3]. The increasing relevance of AI across disciplines and sectors has led to a wide variety of views on the nature of this transformation, but it is most often envisioned through a technological lens: self-contained

AI technologies that change (or replace) the processes we use to work with information. However, the technology-centric perspective that drives much of AI development and innovation fails to capture important aspects of the complex, human contexts in which AI is used and is increasingly insufficient to address the challenges of effectively and ethically using AI in real-world settings [4–6].

The ongoing changes in the AI landscape are not only affecting the scope and scale of AI use: they are also reshaping what it looks like to work with AI technologies in practice. For most of its history, AI has been a specialized, technical discipline, focused on methodological innovation and requiring advanced training and significant computational resources to use effectively. As technologies have matured, AI is in the midst of an expansion from expert technique to everyday toolkit, opening up new opportunities for its use to ask novel questions and perform complex analyses without the need for deep technical expertise or innovation [7,8]. This growth in the purposes for which AI is used has outpaced the professional training and development of best practice required to ensure that AI use is both effective and ethical. This has left significant knowledge gaps in the specific skills and competencies involved in using AI in practice, especially in newer and more interdisciplinary contexts, and the specific skills and competencies needed to work with AI as part of processing information.

Part and parcel of these knowledge gaps, and a key challenge created by the growth of AI application and research, is the fact that different disciplines, fields, application contexts and stakeholders have different conceptualizations of what AI is and what is involved in using it. Each discipline in which AI is studied or used tends to define and operationalize AI in its own ways, with explicitly interdisciplinary approaches to AI in practice being few and far between [9–11]. These different conceptualizations of AI in disciplinary literature and applied contexts are rarely surfaced and often in conflict with one another, functioning as competing epistemologies rather than complementary viewpoints. As AI use becomes more urgently interdisciplinary, by virtue of the diversity of AI users and the breadth of impact from AI applications, we need strategies to help partially reconcile these different understandings of AI in practice and to work effectively across varied perspectives on AI to develop shared practices for AI use.

In this article, I propose a novel framework, which I call *AI Thinking*, which holistically models key decisions and considerations involved in bringing AI into practical use. AI Thinking is a conceptual, competency-based model of the practices involved in applying AI in context, designed to help users of AI systems ensure that their applications are both goal-driven and context-sensitive. To do this, I approach AI as an information *methodology,* rather than an information *technology*, in the process describing and affecting how we work with information as a whole. AI Thinking bridges disciplinary perspectives, illustrating how the design and implementation of AI systems are inextricable from the practical contexts they operate in and how the social systems surrounding AI are reshaped by the process of computation. By providing practice-based connections between different disciplines and practical contexts, AI Thinking encourages users to work with AI as an interconnected set of technical processes and social practices. By framing AI use in terms of the practices that connect specific problems, technologies and contexts, AI Thinking highlights shared concerns and reference points that help connect different definitions and practical understandings of AI. The AI Thinking framework thus makes clearer links between academic disciplines and across diverse contexts of AI use and can help to lower communication costs between different perspectives on what AI means and how to understand, train people in and manage its use.

2. The value of a framework for AI in practice

The nature of working with AI, and the range of activities falling under the 'AI' umbrella, has been changing dramatically. As the availability and ease of use of commercial AI technologies have transformed perceptions of who can be an AI user, and potential value propositions and purposes for AI use have expanded across professional contexts, shared reference points are needed to ground discussions of using and managing AI and building best practice. The AI Thinking framework is intended to provide just such shared reference points for the use of AI in practice, making it easier to join up different ways of thinking about AI and the types of decisions and processes involved in its use.

2.1. Actors in AI ecosystems and users of AI Thinking

The question of who is seen as acting in the space of AI is highly contested, and the boundaries between AI users, developers and those affecting AI use in practice are increasingly blurred [12–14].

By focusing on the processes informing AI in practice, the AI Thinking framework avoids the need to fix specific definitions and categories of AI actors, instead providing points of access, assessment and intervention for anyone involved in a given process. AI Thinking thus adopts an ecosystem orientation to AI [15] in which diverse actors interact with each other and the processes of AI use in multiple ways, and the locus of action to shape the ecosystem is sited at these points of interaction.

AI Thinking is, therefore, relevant to users of AI systems, informing the way they interact with AI systems, choosing inputs and working with outputs; AI developers, shaping the ways in which AI technologies are designed and implemented to work in different contexts; IT and team managers, determining the platforms and organizational practices with which AI use is managed; policymakers (in organizations and government), setting the standards against which AI use will be measured; and data subjects, to illustrate how their data inform the use of AI and its impacts in practice. These roles may often cross over with one another, and with other roles and processes around AI, such as marketing and sales, where AI Thinking can inform clearer communication of how potential users can approach, benefit from and manage risk in AI use.

The AI Thinking framework is not intended to comprehensively address all of the decisions and considerations relevant to each of these roles nor is it envisioned that any one individual will necessarily work with or have the opportunity to shape all of the components of the framework. Rather, by providing a shared understanding and language of a set of key practices, decisions and considerations shaping AI use in practice, AI Thinking enables bridging between the different foci of varying and evolving AI roles and providing points of entry to anyone involved in—or simply seeking to understand—the processes of AI in practice.

2.2. Using the AI Thinking framework

AI Thinking is a multi-purpose descriptive framework that provides a common framing and structure for different aspects of AI use in practice. It is, therefore, relevant to a wide range of contexts surrounding the training and management of AI actors and the production and use of AI systems. In education, AI Thinking can help guide the training of AI professionals, including developers, technical managers and AI researchers in technical and social sciences. In industrial practice, AI Thinking provides a structure for bringing together interdisciplinary AI teams, structuring and managing their collaborative efforts, tailoring AI systems to specific contexts of use and evaluating the efficacy of those systems in practice. In policy and civil society, AI Thinking helps to illustrate the diverse points of assessment and intervention around AI technologies in practice, supporting more effective measurement of AI impacts and more targeted policy interventions and development of best practice. For the wider public, AI Thinking can help to demystify AI and illustrate the ways in which AI in practice is fundamentally shaped by familiar concerns of people and process and not solely the purview of often-inaccessible technological innovation. I do not claim that the AI Thinking framework comprehensively addresses what is needed for each of these purposes; rather, it provides a shared and tangible starting point for building common understanding across these different areas of practice, which can then be supplemented with more context-specific resources.

3. Bridging AI divides across disciplines and contexts

To illustrate the need for a more interdisciplinary conceptualization of how we work with AI in practice, I first outline key differences in the broad range of conceptualizations of AI. The aim of the AI Thinking model is to bridge between competing (and sometimes incompatible) perspectives and approaches across disciplines, including computer science and engineering, science and technology studies, critical data/AI studies and the broad landscape of AI applications. Negotiating the varying conceptualizations adopted in these fields, and the different visions they present of what matters most in studying and managing AI, is one of the most significant current challenges in responsibly applying and managing AI in practice. The competency-based model of AI Thinking presented in this article reflects on these varied understandings together with the growing body of evidence on AI practices in diverse settings, including research and innovation, business, health and policy [16–22]. In the sections below, I briefly outline major differences in AI understandings and the historical contexts from which AI practices and the AI Thinking model have emerged.

3.1. Operational terminology

The current environment of rapid change in AI technologies and applications is also one of the rapid changes in language. The range of disciplinary perspectives addressed in this article makes it necessary to establish an operational terminology with which to describe AI in practice and to set the scope of what aspects of AI technologies and practice are in focus.

3.1.1. AI use

I focus, in this article, on the use of AI for information processing—e.g. transforming and analysing data, learning from evidence and informing action—common across many applications in business, research, policy, healthcare and other decision-making settings. The AI Thinking framework has potential to inform more creative uses of AI, e.g. for content generation or creative practice, and may prove helpful in exploring evolving practice in this area. This focus on information processing does not exclude the use of generative AI, which has significant (and developing) applications for both information processing and creative use (cf. [23,24]).

3.1.2. AI technologies

Artificial intelligence, as a family of technologies, remains notoriously difficult to define. The evolving scope of AI technologies—i.e. specific models, algorithms and technological products—includes interactive systems such as generative AI, foundation models for adapting general-purpose knowledge and more specialized AI systems (sometimes referred to as 'narrow' AI) targeting specific applications through bespoke machine learning or expert systems [25,26]. The use of 'AI' in this article does not distinguish between these types of AI technologies but focuses on common processes and ways of thinking that inform all of these types of AI in practice.

3.1.3. AI systems

AI use is often discussed in terms of 'AI systems', but the nature and boundaries of these systems have varied definitions. In many cases, for example in industrial practice [27] or international policy contexts [26], AI systems are defined in terms of software and technical systems alone. In presenting the framework of AI Thinking, I follow thinking in critical data studies by taking a wider view of the contexts in which AI software and technologies are inextricably situated and recognizing the essential role of non-technical stakeholders and perspectives in setting the shape and the purpose of AI technologies [28]. Attending to the broader milieu in which AI operates calls for a systems thinking approach [29] and reflects the inherently sociotechnical nature of AI [6]. Thus, from an AI Thinking perspective, an 'AI system' includes not only the software and technologies through which AI computation occurs but also the broader system of people, data and processes that motivate and make use of AI computation.

3.2. Differences in AI understandings across disciplines

The historical contexts of different fields and areas of AI application have naturally led to different conceptualizations of the same ideas. These differences, which are not always immediately visible, hinder the interdisciplinary dialogue and collaboration necessary to make AI use ethical, explainable and effective in practice [7,30,31]. Differences especially in the language around AI and the scope of who is involved in its application can also reflect deeper underlying differences in how AI is conceptualized between disciplines and contexts, and what aspects of AI systems are considered most important.

In my proposed AI Thinking framework, I aim to provide points of common reference and comparison to work across different understandings of AI by joining up technical and social practices involved in AI use. Here, I briefly outline three common ways of framing AI observed across different disciplinary contexts: AI as a linear, cyclical and relational system. I show how an AI Thinking approach can help to bridge their distinct perspectives on data, computation and the scope of AI systems.

A *linear* understanding of AI, commonly encountered in computer science and engineering disciplines, focuses on the process of computation: beginning with input data, performing a sequence

of transformations and analyses and producing output for further use. On this framing, computation is emphasized and data are pre-existing, material objects to be consumed, transformed and/or generated. AI systems are typically considered through a technological lens, as separate concerns from the production of data or use of AI outputs. This perspective is reflected in the directional flow diagrams commonly used to illustrate AI model structures and data analysis in computing literature (cf. [32,33]) and some applications [34].

A *cyclical* understanding of AI, commonly seen in AI applications such as in healthcare or finance, takes a more process-oriented view of AI use, including data collection, analysis, interpretation and assessment, and action taken in response, which frequently produces new data for a subsequent cycle. On this framing, computation is one piece of a larger process which forms the broader context for AI systems as embedded technologies. Data are emphasized, as representative of the information being moved, understood and acted upon in the overarching process. This perspective is reflected in visualizations and discussion of AI systems in professional practice [35,36] and in the growing literature on human–AI teaming [37,38].

A *relational* understanding of AI, more commonly encountered in social sciences and humanities disciplines, focuses on the people, perspectives and purposes behind the use of AI in sociotechnical contexts. A wide range of theoretical approaches are used to investigate AI from more relational perspectives, including systems thinking [29] and assemblage theory [39], as well as more specific models such as data journeys [40]. On this framing, computation is frequently a means rather than a process, emphasizing what Amoore describes as 'the capacity to generate an actionable output from a set of attributes' [41], and the data that describe these attributes are a dynamic and contested site of practice. This perspective is frequently reflected in critical studies of AI and analyses of AI harms [42,43]. The definition of AI systems adopted for this article most closely aligns with a relational understanding of AI.

To create a cohesive model for AI use across disciplines, these differing understandings of AI must be treated as what they are: equally necessary and informative for a full picture of AI Thinking. The *linear* framing is essential for choosing between and working with specific AI technologies, which are engineered to reflect a linear pathway from input to output. At the same time, the *cyclical* framing is key to assessing data sources and their implications for stakeholders in the processes where AI is brought to bear. And a *relational* perspective is required to understand the contextual motivators and impacts of AI use, and how success and risk can be most meaningfully measured.

The AI Thinking framework reflects elements of each of these framings, helping to reveal interconnections between them in practice and to find common ground within the interdisciplinary teams necessary for AI research, application and regulation. AI Thinking thus helps to join up aspects of AI systems: illustrating how the relational networks and data cultures affect the input data used for linear computation, and interpret and use the output of that computation; and how the cyclical nature of pre-existing processes helps to shape the linear process implemented in the design of an AI technology for practical use. By understanding that these conceptualizations of AI are not in competition, but rather reflect different and equally necessary parts of the same sociotechnical systems, AI Thinking can help begin to create a cohesive model of AI use across disciplines.

3.3. Roots of AI Thinking: specialization and practice

As well as drawing on a variety of disciplinary conceptualizations of AI, the AI Thinking framework explicitly builds on the practices of statistical thinking developed in the late nineteenth and early twentieth centuries. Similar to the way statistical methodologies matured and gave rise to wider statistical practices at the turn of the twentieth century, AI Thinking is rooted in the maturation of the AI field and its expansion from a technical specialism into a broader methodology.

Early development and use of statistical methods were primarily a matter of mathematical research, though often motivated by and entwined with the needs of states to understand changing populations [44]. As statistical methods matured, however, they began to have significant value for other disciplines grappling with measuring and understanding complex phenomena, and by the mid twentieth century statistical methodologies had grown far beyond the original bounds of statistics as a specialized discipline and become essential to research and practice in psychology, medicine and the natural sciences [45]. In the later twentieth and twenty-first centuries, statistical methods have become increasingly essential to decision-making in business also [46]. While statistics has certainly continued to develop as a rich discipline and its own specialism, fundamentals of *statistical thinking*— considerations of sample size, recognition of random effects, uncertainty as part and parcel of research

[47]—have become core competencies for contemporary research and practice across disciplines and a wide variety of applied contexts [48].

AI methods are beginning to play a similar role, providing new tools for working with the increasingly complex combinations of data and knowledge sources that characterize contemporary information systems. AI techniques, particularly machine learning, are reshaping our ability to learn from multiple information sources and combine data in complex ways. For example, in research spheres, the rapid diffusion of AI methods across research practice, from combining exascale imaging data [49] and protein folding [50] to supporting data analysis in clinical trials [51] and digital humanities [52], is moving us towards a research ecosystem where using AI methods is necessary for competitiveness in many areas of inquiry. Similar transformations are occurring in business, healthcare, policy and other areas of information processing [19,53–55]. Thus, while AI research as its own specialism will also continue to mature, fundamentals of AI Thinking are needed to help inform why, when and how to use AI methods effectively and to understand both their limitations and strengths.

Like statistical thinking, AI Thinking is rooted in the professional practices and knowledge that are the foundation of training for experts in the field. Statisticians learn how to assess sample size, measure uncertainty, characterize populations and so forth, and these are key building blocks with which deeper expert knowledge is built [47]. Similarly, AI experts learn fundamental skills such as problem formulation (i.e. mapping a particular application to a known type of task with established methodologies), selecting and managing data, choosing models and algorithms, and interpreting the results of AI computation. But at this point, practices diverge. Expert statisticians develop deep knowledge of mechanics and principles and may innovate with new measures and rich characterizations of complex phenomena; AI experts develop deep knowledge of mathematical and cognitive theory, and may innovate complex new model structures and learning algorithms. But for most users of these methodologies, this depth of expertise is unnecessary: the key knowledge is *operational*, knowing how to develop an appropriate process to ensure reliable evidence or synthesize information. This operational knowledge can then be combined with a deep understanding of the context in which the methods are applied, whether in specialized research, policy, decision-making or other kinds of information processing. This bifurcation into *technical expertise* on the one hand and *operational methodology* on the other has come to characterize the dual—though actively contested—nature of statistics in practice [56] and is a productive framing for approaching the emerging dual nature of AI.

It is important to note that AI Thinking is not the next evolution of statistical thinking nor does it supplant statistical thinking. Indeed, as most contemporary AI relies on statistical methodologies, statistical competence is essential to AI competence and helps AI users to cast a critical eye on the uncertainty inherent in using AI methods. AI Thinking and statistical thinking thus sit alongside one another as fundamental competencies shared across areas of working with information.

4. AI Thinking: an interdisciplinary, competency-based framework for AI use in practice

AI Thinking is a conceptual, competency-based model of the practices, decisions and considerations involved in applying AI systems in context. AI Thinking is process-oriented and practice-based; competency models thus offer a natural fit for defining what AI Thinking looks like, and a practical foundation for teaching AI Thinking as a foundational information skill across disciplines and contexts [57]. I outline below five core competencies of AI Thinking, representing key skills for working with AI methods as a practical information processing tool. As AI systems are typically a team effort, it is not necessary that any one person master all five of these competencies; rather, they offer a way for identifying and seeking out the skills needed in an AI team. These competencies present a starting point for further expansion into a systematic AI Thinking framework, akin to the statistical thinking model of Wild & Pfannkuch [47], and drawing on early development of professional competency frameworks for AI [53,58].

The AI Thinking framework, illustrated in figure 1, comprises five major elements of working with AI methods in practice: (i) initial alignment of AI use with *process*-oriented needs, (ii) the *formulation* of AI approaches to respond to those needs, (iii) assessing and selecting potential AI *tools and technologies*, to implement the chosen formulation, (iv) assessing and selecting *data* sources to inform the chosen tools and technologies; and (v) ensuring the full lifecycle of AI system design, implementation and management is rooted in the *context* of use. Figure 2 summarizes key elements of
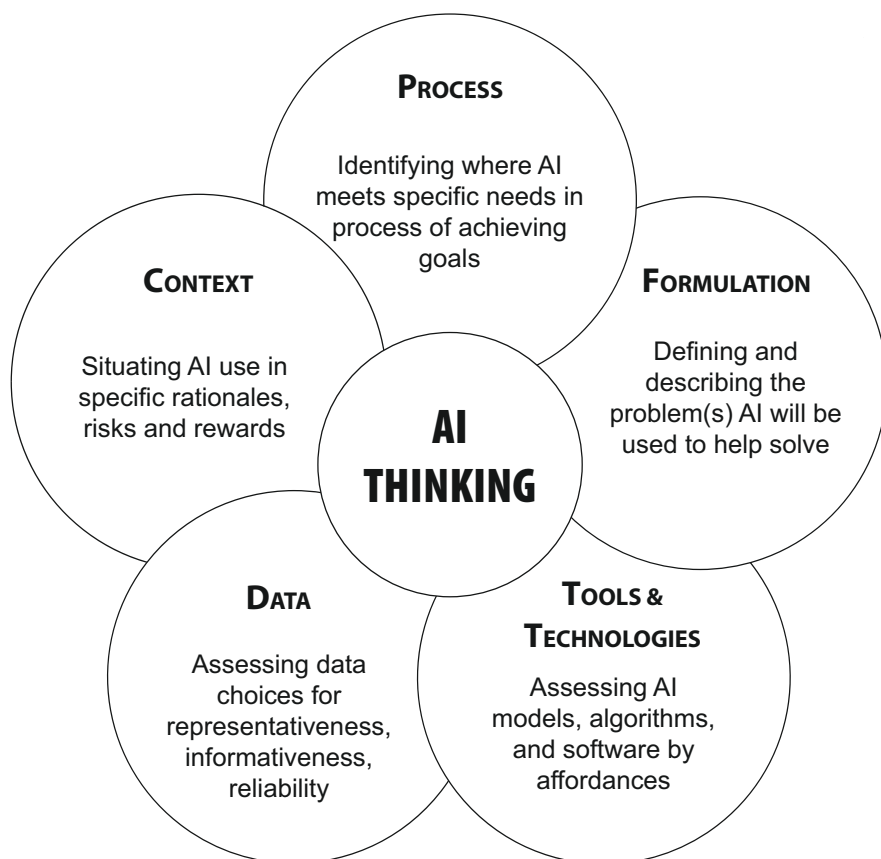
Figure 1. Five key competencies of AI Thinking. These competencies are arranged in a circle to show that they are interrelated and used in parallel rather than separate concerns addressed in sequence.

each of these five competencies and illustrates their interrelationships within a linearized process of AI system design.

4.1. Process: motivating AI use by specific needs

Definitions of AI have changed as technologies have evolved, but AI research and innovation have been consistently characterized by a focus on *doing specific tasks*, even as AI technologies themselves have become more general-purpose [26]. However, adopting AI methods in practice is often motivated by a pressure to be innovation-led and stay ahead of the technology curve rather than as a response to specific needs for information processing that AI methods can help address [59]. The greatest practical benefits from AI have historically been achieved when its use is motivated by specific needs in the service of broader goals: e.g. development of AlphaFold for identifying candidate protein structures from a vast possibility space [50]; now-ubiquitous speech recognition technologies initially developed for supporting call centres [60]; and automated analysis of medical imaging to support cancer diagnosis [61]. In each case, a particular part of the process was identified for AI intervention, and AI use was guided by an understanding of that process.

Adopting a process-oriented motivation for AI use, and ensuring that AI applications are goal-driven rather than innovation-led, is the first key competency of AI Thinking. Taking a process-oriented approach to AI involves three distinct elements:

— AI use is *goal-driven:* developing a new AI technology or bringing an existing technology to bear is motivated by a desire to make use of information in service of a specific goal.
— AI use has a *defined scope* within the process of achieving that goal, typically in terms of specific steps or distinct operations needed. There may be multiple distinct scopes for AI use within a process, which may be approached individually or jointly.
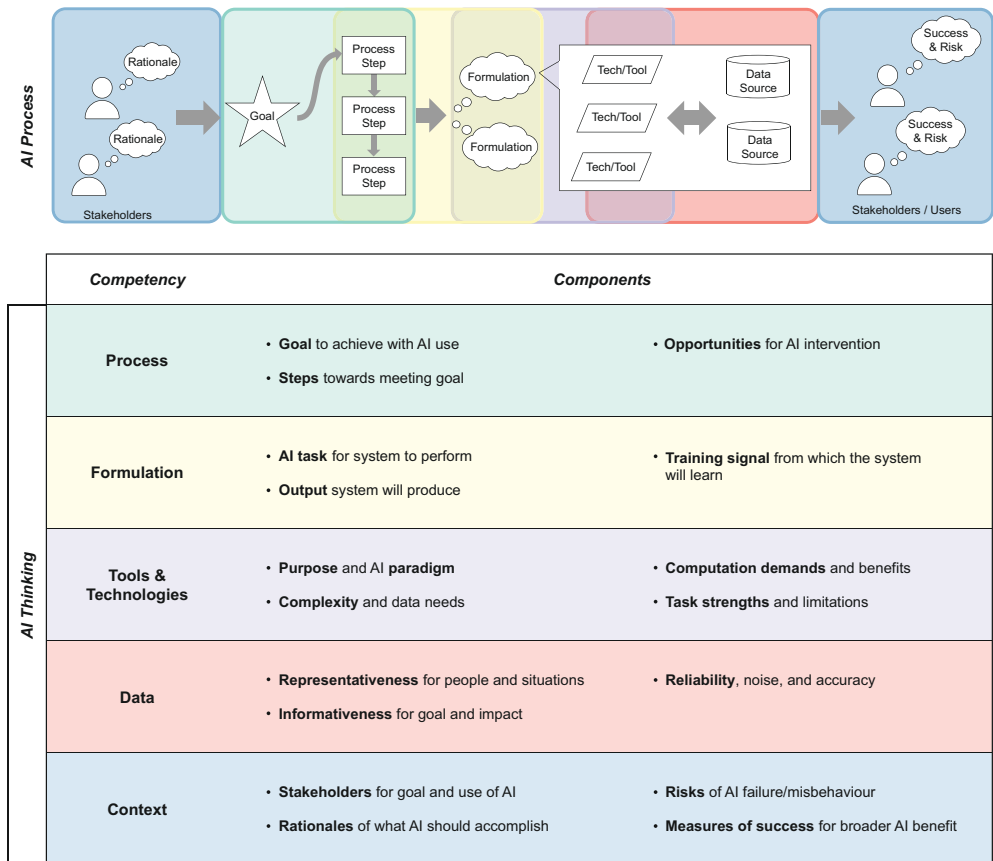
Figure 2. The competencies of the AI Thinking model. Each competency is aligned with one or more portions of a linearized Illustration of the process of using AI.

— AI systems are designed to respond to *specific opportunities* arising within each scope for use, in terms of specific information availability and information needs at each step. Not all steps of a process will present appropriate opportunities for AI use.

This process-based approach does not preclude opportunities for innovation and discovery from using AI in a more 'end-to-end' fashion to complete multiple steps or the entire process at once (discussed further below), but approaching AI use from this foundation in process clarifies opportunities and scope for intervention, provides multiple points for innovation as well as risk mitigation and ensures that AI use is guided directly by desired goals. This approach of breaking down goals into distinct steps for AI intervention is particularly valuable when information processes need to demonstrably adhere to community norms and standards, as in research practice or business processes, or even by statutory or regulatory requirements.

Beginning from a foundation in process, AI use in practice is then informed by the next AI Thinking competency, describing and formulating the specific problem that AI use is intended to help solve.

4.2. Formulation: describing the problem to solve with AI

AI is a discipline of problem-solving, and identifying how best to describe the problems to be solved has been a core component of AI development and use from its earliest days [62]. *Problem formulation* involves creating formal descriptions of AI problems (typically what type of output should be produced from some type of input) and is an essential step in bridging from the complexity of real-world problems to the well-defined domain of mathematical models in AI systems. AI problem formulations have evolved from stricter mathematical design under planning paradigms of AI [63,64] to more operational definitions of tasks to perform and ways to conceptualize AI computation in current machine learning paradigms [65], but the core step of formalizing the operation of an AI system remains essential for AI development and evaluation.

These shared definitions are powerful tools, making it possible to re-use and repurpose a common base of AI models and methods. For example, approaching a new, complex text labelling problem as a named entity recognition problem (in which text segments describing real-world referents are identified) enables repurposing a wide existing base of technical methods for a new purpose [66]. However, as a variety of recent critical analyses have illustrated [67–69], problem formulation is highly epistemically loaded, involving normative decisions about what matters in the operation of an AI system and what that system can meaningfully say about its inputs and the world. Negotiating the tension between the benefits of shared problem formulations and the epistemic positions they represent is thus an essential part of AI use in practical contexts.

The second key competency of AI Thinking is thus understanding and negotiating the process of problem formulation for AI systems in practice. This comprises three main elements:

— The *AI task* the system is to perform, such as image classification, text generation or speech recognition. This defines the templates by which the system should operate, and opens up opportunities for using some methods and models while closing the door on others.
— The *output* the system will produce: will it be one categorical label? Multiple numeric scores? A sequence of text? This affects how the AI system can inform its users, and what it is capable of saying about its inputs.
— If using a machine learning approach, the *training signal* from which the system will learn. This includes both what will be used to signal the desired or undesired outcome of AI processing, and the measures used to penalize errors and fit the AI model to the observed data.

Any practical information processing need where there is appropriate scope for AI intervention is likely to be open to multiple alternative problem formulations and may in fact consist of multiple related sub-problems. This space of possible formulations and decomposition of sub-problems is illustrated by a wide range of examples of practical AI applications: Cheng *et al*. decompose planning-based models for business decision-making into multiple related decision processes for greater clarity and optimization [70]; Newman-Griffis & Fosler-Lussier compare AI paradigms for categorizing functional status information [71] and Isagah & Ben Dhaou describe several diverse examples of AI problem formulations in government [72].

Once a process-oriented goal for AI use has been transformed into specific problem formulations, the third AI Thinking competency addresses using the affordances offered by different AI tools and technologies to select appropriate methodologies.

4.3. Tools and technologies: assessing AI affordances

As AI technologies advance at a rapid pace, keeping track of the changing landscape can seem unmanageable, and AI users often feel like new skills must be constantly relearned with each new advance [73]. Comparing and contrasting available technologies for a given use of AI, and adapting and responding to new technological advancements, are thus key skills for working with AI in practice and building sustainable AI systems that can weather the storms of technological change.

The third key competency of AI Thinking is to conceptualize AI tools and technologies in terms of the *technological affordances* they present, and how these affordances align with the intended use of AI. Technology affordances [74] extend psychological affordance theory to describe the perceived possibilities that a given technology offers to potential users, and the strengths and weaknesses it presents for different possible uses. While any specific technology may be repurposed, hacked or extended to do a wide variety of things beyond its original intended purpose, affordances are a powerful way of characterizing likely and easily perceptible opportunities a technology offers, which form a core functionality for practice. Focusing on affordances, and how they align to specific information processing needs, thus provides a clear, process-oriented strategy for comparing available AI technologies and for assessing new technologies—and the risk and reward in their adoption—when they emerge [75].

Each AI tool and technology is engineered with a particular task in mind, and each offers specific perceptible affordances for information processing. These affordances are dynamic, as new uses are discovered and reshaped, and may go far beyond the original motivating task; for example, the growth of prompt-based learning in generative AI [35] or the shift in use of AlexNet [33] and other benchmark computer vision models from image classification to broader representation of image features for many purposes [76]. Assessing the affordances of AI technologies involves four key components:

— The *purpose and AI paradigm* for which a given technology was designed, including its connections to specific problem formulations as well as identifying where overlapping or complementary task definitions provide opportunities (and limitations) for interoperability and reuse.

— AI technologies have variable *complexity and data requirements* that affect their viability and effectiveness for different applications. As a general rule, more complex AI models have greater capacity to capture complex patterns and relationships but also require larger volumes of data to work effectively. Task complexity and data availability vary widely across contexts and are essential for selecting appropriate technologies.

— AI technologies must also be assessed in terms of their *computational requirements,* i.e. the hardware required to execute model calculations efficiently. While the shift to cloud computing has reduced the need to manage on-premise hardware, financial cost and environmental impact of different AI technology options are directly driven by computing requirements.

— Each technology, and each task formulation underpinning it, presents *strengths and limitations* for performing different tasks. These may be characterized by the initial research and development of new models, emerge from later, evolving uses of those models, or be considered as adaptations to an underlying framework.

Rather than assuming that newer or higher-performing technologies are necessarily better, assessing the affordances of AI technologies along these dimensions provides a balanced way to select the appropriate tool or technology for a given application. For example, in detecting and modelling rare diseases, a simpler model may be preferable to data-hungry deep learning [77], while the reverse is true for cancer detection and prognosis, where complex relationships between hundreds of variables are needed and rich multi-modal data are available to learn from [78]. A large language model (LLM) may be an appropriate choice for processing common financial reports [79], while more bespoke analysis may be a better fit for analysing individual forms or detailed patterns of language use [80]. As new technologies emerge, the same process of assessing their affordances can provide a clear pathway to determining whether they should be adopted over current approaches and identifying what new problems they may be fit to tackle.

With a process in mind and an understanding of the affordances on offer from different AI tools and technologies to support that process, the next key competency of AI Thinking is identifying the appropriate data sources to guide AI implementation for the desired goal.

4.4. Data: informing AI use with appropriate information

Data are often said to be the foundation of AI systems [20]. What is less often noted is that data are also a choice, both in the process of their production and in their use in AI. The idea of 'raw data', unprocessed and therefore neutral, lingers in technical discourses [27], but extensive scholarship has shown how the production of data reflects multiple choices about what counts, what to collect and what to record about a given subject [81–83]. There are also practical concerns of the compatibility of different data sources and datasets with the available tools and technologies, reflected in the well-established central role of choosing appropriate training data in machine learning research [84]. In a process-oriented AI Thinking approach, the fourth key competency is thus to assess which of the available data fields, sets and sources are most appropriate to inform the desired goals and the processes in which AI is to be used.

Assessing data sources is thus a matter of both epistemology and practicality, with each having significant power to affect bias and fairness of data and the ethical practices they reflect and act on [81]. Different data choices will embed different distinctions and perspectives on the task at hand and will work differently with available tools and technologies. While there are as many ways to assess data as there are data points, practice-based assessment of data begins with three key criteria of representativeness, informativeness and reliability:

— *Representativeness* has many (and contested) meanings [85] but can be broadly considered in terms of data's ability to reflect the full diversity of the people or situations they are meant to convey information about.

— *Informativeness* can be distinguished from representativeness, in terms of how well data are aligned with the desired goal for AI and how clearly they evince useful learning or patterns for that goal [86].

— *Reliability* includes noisiness as well as the likelihood of data being collected consistently and accurately in practice [87,88].

Each of these criteria may be balanced differently for different situations and goals. For example, foundation models are built on the principle of an extraordinary volume of data being sufficiently informative and representative to capture diverse general patterns despite lower reliability; in scientific analysis, data reliability may be paramount; in exploring alternative decisions a smaller set of diverse and informative data may be more valuable than data representing population norms. In addition, practical AI applications may involve multiple entry points for data with distinct balances between these criteria: e.g. a diagnostic model in healthcare that begins with a foundation model and is progressively fine-tuned with smaller datasets which may be less representative of human diversity but are more informative for the diagnostic goal.

When exploring AI tools and technologies as options for intervening in a specific process, it is thus essential to assess the appropriateness and fit of different data sources for the task at hand. The final key competency of AI Thinking ties these three elements of process, tools and data for using AI together in the specific contexts where AI is used in practice.

4.5. Context: shaping AI use, benefits and risks in situated practice

Using AI is, at its heart, about making use of computation to help us learn something, produce something or do something in the world. As we have seen, the processes in which we bring AI to bear, the tools and technologies we choose to use and the data sources we draw from each carry with them pieces of a broader picture in which AI use is only one element moving between different goals, actors and perspectives. When this broader picture is ignored, or left as a separate concern from the technical design and use of AI systems, those systems inevitably miss the mark [89] and may create substantive harm [90]. Recognizing the contexts in which AI use is always situated, and understanding the implications of those contexts for AI's effectiveness, benefits and risks is thus the fifth key competency of AI Thinking.

There are, of course, many contextual aspects that bear on AI use, and analysing the sociotechnical assemblages of AI in practice is the subject of a growing research literature (cf. [91–93]). Throughout the design, implementation and management of bringing AI methodologies into practice, a strong foundation for assessing AI contexts begins with:

— The *stakeholders* for the proposed use of AI. This includes those who set the initial goal for AI use to support, as well as those who produce the data used by AI systems and make use in turn of the outputs those systems produce.
— The *rationales* these different stakeholders have for using the AI system and what it is meant to accomplish. For example, a senior leader may look to AI use for efficiency gains, while a researcher or business analyst looks to the same system to produce a specific insight from the complex data they work with.
— The *risks* posed by different points of failure or undesired behaviour in the AI system, with respect to these rationales. For example, an LLM hallucination or a spurious relationship from a machine learning model will have different impacts on an AI user, who may be concerned with accuracy in a specific instance than on someone affected by an AI-informed decision process, for whom recourse and minimizing harm are paramount.
— The *measures of success* for stakeholders in the AI system. Task-based performance metrics assess how well an AI technology performs its stated role but may not capture the impact of that technology on someone using it as one piece among many, for example in a decision-making process.

These components of AI context may seem distant from the process of developing and implementing AI tools and technologies, but they are essential to connecting that process to the people and processes in which AI will actually be used. AI Thinking, as a model for bringing AI methodologies into practice, thus aims to connect the dots between technology, the people who use it and the purposes it is meant to serve.

5. Case study of applying AI Thinking in practice

To illustrate the role of the AI Thinking framework in practice, I outline a hypothetical case study of using AI for processing information about health and well-being, an area of practice where the rich tradition of medical ethics highlights gaps in current approaches to AI ethics [94]. The case study examines the use of AI methods as part of prioritization of patients for organ transplants, and illustrates how AI application might unfold with and without AI Thinking.

5.1. Scenario

Prioritizing potential recipients for organ donation, and ensuring successful transplantation, are highly time-sensitive and information-intensive processes. The use of rich data sources and AI systems to help tackle these challenges has been growing, and we will take as our first illustration of putting AI Thinking into practice a recent paper using machine learning-based simulations of transplant outcomes to inform decision-making [95]. We will imagine an extension of this methodology, implemented for prioritizing potential transplant recipients in a large health system serving a demographically diverse, primarily rural population. We will consider the practices of the health informaticians building the AI system, the healthcare professionals using it to inform transplants, the patients receiving transplants and the health system administrators assessing service outcomes and quality.

5.2. Without AI Thinking

Health informaticians design the AI system to tackle the motivating need directly and rank potential recipients by outputting a single score based on an overall simulation of post-transplant health outcomes. The score is calculated based on the simulated outcome of the transplant, which is based in turn on a combination of demographic and health data, including records of past healthcare encounters. The AI technology is implemented using a single central deep neural network, which is trained based on a large volume of data sourced from the implementing health system as well as reference data from a large urban system elsewhere in the country. This is done to provide the greatest amount of data to learn from, in order to improve the model's robustness and sensitivity to variation. The model's performance is measured by its fidelity to past transplantation decisions.

Most healthcare professionals take the output score of the AI system and combine it with other key information, such as specific relevant laboratory results, distance to the patient, and size of the organ relative to the patient, to decide who to call for transplant. Some professionals mistrust the AI system and manually review health data to make recommendations, frequently arriving at different priority scores than the AI technology. Health system administrators measure success of the process using common service metrics such as 30 day mortality and rehospitalization rate. A periodic audit after a year of using the AI system reveals that patients from urban areas are recommended for transplants at a higher rate than patients from rural areas, and that racial disparities in health outcomes are more pronounced for rural patients. These differences are consistent across professionals who use the AI system, and less pronounced for some who avoid it while being more pronounced for others. Patients and families are left with limited information about why they were not selected for transplant or why outcomes were worse than expected, with minimal transparency into the decision process or recourse to contest it.

5.3. With AI Thinking

Re-examining the situation with the AI Thinking framework helps to bridge gaps between understandings and practices in this scenario and to reduce the racial and geographic disparities observed in health outcomes.

5.3.1. Process

Prioritizing transplant recipients is the overall goal, but this is implemented in professional practice by combining a number of standard key variables with an overall understanding of a patient's health status [96]. Through discussion between the health informaticians and healthcare professionals, the AI technology is instead designed to target the more specific scope of assigning a transplant score to the

patient's health status specifically. The other variables used in the ranking process are clearly defined by clinically motivated rules, and not considered targets for AI intervention.

5.3.2. Formulation

The need for transparency and accountability for health system administrators, and for insight to support clinical discretion for healthcare professionals, means that rather than a single atomic score the task is formulated as simulating multiple component health outcomes, represented as clinically motivated categories or continuous scores. These component outcomes are then transparently combined to an overall score by an adjustable formula. This system is trained based on the final ranking, which combines the AI-produced health score with the other, rule-based variables used by the clinicians, to model its use in practice.

5.3.3. Tools and technologies

Rather than a single, large deep neural network, which imposes significant data and computation requirements for the health system, the component health outcome models are implemented using a mixture of support vector machines, random forests and smaller, targeted neural network models, with significantly lower technical requirements and greater opportunity for accountability and explainability in the modelling process.

5.3.4. Data

A review of the available data, motivated by healthcare professionals' observations, shows that a significant fraction of rural patients accessing the health system do so only through visits to the emergency department [97], and that structural racial bias emerging from the system's historical context has produced unequal delivery of services to some racial groups. To reduce the structural bias embedded into the AI system, the interdisciplinary team works together to avoid methods that assume data are race neutral [98] and minimize reliance on data that are systematically missing for rural patients.

5.3.5. Context

Input from all stakeholders—informaticians, healthcare professionals, health system administrators and patients—is incorporated throughout the application of all other competencies. System performance and outcomes are measured by monitoring mortality and rehospitalization rates for transplant patients and explicitly measuring population disparities. The additional transparency created in the system by a process-oriented approach enables more proactive and constructive dialogue with patients and families who are affected by the system decisions.

5.4. Summary

In this hypothetical scenario, the AI Thinking framework provides tangible opportunities to bring together the interdisciplinary perspectives and experiences in the AI team in question. The decisions and considerations represented in the AI Thinking model clearly identify points of reflection and intervention to bridge disciplinary divides and help to capture risks of inequitable outcomes from the system that were missed when not joining up stakeholders and disciplines. While this example might well play out differently in practice, it helps to illustrate the role of AI Thinking in working across disciplinary differences to bring AI into more thoughtful interdisciplinary use in practice.

6. AI Thinking in the broader AI discourse

Ongoing rapid transformations in AI have led to many different ways of framing important questions: what AI can (and should) do and why, who should use AI and how, and what is required to achieve real benefits with AI while managing its risks. This section situates AI Thinking within this broader

discourse, with respect to three major reference points: AI literacy, end-to-end approaches to AI technologies, and 'AI-driven' innovation.

6.1. AI literacy and AI Thinking

The need for broader AI literacy is well recognized, and strategies for developing AI literacy in students, professionals and the public are rapidly evolving [99–101]. AI Thinking, as a model for framing AI use in practice, complements current AI literacy efforts in two ways. AI literacy discussions focus primarily on formal education, whether in schools or higher education [102–104]; AI Thinking provides a model to guide training and self-learning in professional and practice-based settings. AI literacy also focuses primarily on understanding how AI works and what its implications are, either from others' use or from one's own use of AI tools produced by others [101,105]; AI Thinking describes the competencies required to use AI methodologies as one piece among many in the context of broader evidence-based practice. AI Thinking may be considered as an element of AI literacy, at least for practice-based audiences, but its focus on AI practice and its methodological, context-based perspective is important to distinguish from more general-purpose AI literacy efforts.

6.2. A process model versus 'end-to-end' innovation

As 15 years of deep learning advances have made more complex tasks more achievable for machine learning, the movement towards 'end-to-end' AI approaches has continued to pick up steam. End-to-end AI treats a sequence of related tasks, each of which may be simple or difficult on its own, as a single, highly complex problem, such as in drug discovery [106], speech recognition [107] or even autonomous driving [108]; this allows (it is argued) for greater innovation and discovery of unantici-pated relationships and approaches to solving difficult problems. AI Thinking, as a process-oriented approach involving decomposing goals into multiple specific points for AI intervention, seems at first to run counter to the end-to-end model. However, a process-based orientation does not necessitate a process-based intervention: with appropriate AI tools and technologies and supporting data sources available, an end-to-end approach may well be supported under the AI Thinking model. What AI Thinking provides is a clearer framework for identifying whether an end-to-end or a step-by-step approach is most appropriate to the specific goal, tools, data and context at hand. Process-based AI Thinking can also help to identify clearer points for evaluation, control and risk mitigation in complex AI systems, helping to tackle some of the longstanding problems in end-to-end approaches [109].

6.3. AI Thinking or AI doing? Being AI-informed rather than AI-driven

The idea of 'AI-driven' innovation has gained common currency in recent years, in diverse settings including policy [110], science [111] and business [55]. While the meaning of the term varies in practice, an 'AI-driven' framing creates a persistent narrative of innovation and discovery being *led* directly by the use of AI technologies, as opposed to AI *supporting* the process as one tool among many. The framing of 'AI Thinking', rather than a more action-oriented 'AI doing', emphasizes the role of AI as a support tool and reflects the purpose of the model to describe the design considerations and decisions made in setting up the use of AI, before that informed use then leads to specific action. AI Thinking is thus more oriented towards *AI-informed* design and use, in which AI functions as part of a larger sociotechnical system [112].

7. Conclusion

AI methodologies are transforming the way we work with information. A wide range of disciplines study, use and develop AI and its transformations, but the different perspectives on AI and its use adopted by these disciplines are often in conflict. New approaches are needed to effectively combine the practical insights and analytic methods offered by different disciplines, and to build more holistic interdisciplinary practices informing AI applications in context.

In this article, I have proposed a novel conceptual framework called *AI Thinking*, which models key decisions and considerations to help bridge disciplinary perspectives on AI in practice. Just as

statistical thinking transformed our use and understanding of growing information in the nineteenth and twentieth centuries [113], so AI Thinking can help transform information processing in the twenty-first century. I have presented a competency-based model of AI Thinking, representing the key practices involved in applying AI in context. These five competencies—orienting AI use in process, formulating AI implementations, assessing the affordances of AI tools and technologies, selecting appropriate data sources and situating AI use in context—provide a starting point for rethinking AI skills and training to facilitate more interdisciplinary application of AI, as well as bridging disciplinary perspectives on how AI systems can be better designed, assessed and managed. The AI Thinking framework has relevance to a wide variety of actors in the AI space whose actions inform or are affected by the use of AI in practice, including users, developers, managers and people whose data are the subject of AI analysis. The framework provides a shared reference point to build common understanding and language among these diverse actors and to inform efforts in AI education, deployment and management, analysis and policy.

As AI Thinking is brought more systematically into practice, and AI use continues to become more interdisciplinary in nature, the initial competencies presented here will be further tested and expanded. Wider empirical study of AI practices and needs will be essential to further develop the AI Thinking model and to keep it aligned with evolving AI practice. As the AI discourse continues to grow and change, the relationships between AI Thinking, AI literacy and other key framings in AI innovation and practice outlined here will further develop, enabling new connections and comparisons to inform richer conceptualizations of AI in practice. Strengthening dialogue and collaboration on AI practices across disciplines is vital to achieving practical benefits from AI use while reducing and managing its risks. By bridging conceptual and disciplinary divides in AI research and practice, the AI Thinking framework aims to help shape a more purpose-driven, effective and manageable future of AI in practice.

References

1. Kulkov I, Kulkova J, Rohrbeck R, Menvielle L, Kaartemo V, Makkonen H. 2023 Artificial intelligence - driven sustainable development: examining organizational, technical, and processing approaches to achieving global goals. *Sustain. Dev.* **32**, 2253–2267. (doi:10.1002/sd.2773)

2. Makridakis S. 2017 The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures* **90**, 46–60. (doi:10.1016/j.futures.2017.03.006)

3. Schwab K. 2017 *The fourth industrial revolution*. New York, NY: Crown Publishing Group.

16

4. Bommasani R *et al*. 2022 On the opportunities and risks of foundation models. *arXiv* https://arxiv.org/abs/2108.07258

5. Jarrahi MH, Lutz C, Boyd K, Oesterlund C, Willis M. 2023 Artificial intelligence in the work context. *J. Assoc. Inf. Sci. Technol*. **74**, 303–310. (doi:10.1002/asi.24730)

6. Sartori L, Theodorou A. 2022 A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics Inf. Technol*. **24**, 24. (doi:10.1007/s10676-022-09624-3)

7. Dwivedi YK *et al*. 2021 Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manag*. **57**, 101994. (doi:10.1016/j.ijinfomgt.2019.08.002)

8. Jan Z, Ahamed F, Mayer W, Patel N, Grossmann G, Stumptner M, Kuusk A. 2023 Artificial intelligence for industry 4.0: systematic review of applications, challenges, and opportunities. *Expert Syst. Appl*. **216**, 119456. (doi:10.1016/j.eswa.2022.119456)

9. Kusters R *et al*. 2020 Interdisciplinary research in artificial intelligence: challenges and opportunities. *Front. Big Data* **3**, 577974. (doi:10.3389/fdata.2020.577974)

10. Mazarakis A, Bernhard-Skala C, Braun M, Peters I. 2023 What is critical for human-centered AI at work? – Toward an interdisciplinary theory. *Front. Artif. Intell*. **6**, 1257057. (doi:10.3389/frai.2023.1257057)

11. Newman-Griffis D, Lehman JF, Rosé C, Hochheiser H. 2021 Translational NLP: a new paradigm and general principles for natural language processing research. In *Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics* (eds K Toutanova, A Rumshisky, L Zettlemoyer, D Hakkani-Tur, I Beltagy, S Bethard). Stroudsburg, PA, USA: Association for Computational Linguistics. (doi:10.18653/v1/2021.naacl-main.325)

12. Naik N *et al*. 2022 Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? *Front. Surg*. **9**, 862322. (doi:10.3389/fsurg.2022.862322)

13. Orr W, Davis JL. 2020 Attributions of ethical responsibility by Artificial Intelligence practitioners. *Inf. Commun. Soc*. **23**, 719–735. (doi:10.1080/1369118x.2020.1713842)

14. Kaspersen A, Leins K. 2023 Who decides who decides: what conversations are allowed about artificial intelligence? Australian Quarterly **94**, 9–14. https://www.jstor.org/stable/27203534

15. Stahl BC. 2022 Responsible innovation ecosystems: ethical implications of the application of the ecosystem concept to artificial intelligence. *Int. J. Inf. Manag*. **62**, 102441. (doi:10.1016/j.ijinfomgt.2021.102441)

16. Chui M, Manyika J, Miremadi M. 2018 *What AI can and can't do (yet) for your business*. *McKinsey Quarterly*. See https://www.mckinsey.com/capabilities/quantumblack/our-insights/what-ai-can-and-cant-do-yet-for-your-business (accessed 29 July 2024).

17. Kanbach DK, Heiduk L, Blueher G, Schreiter M, Lahmann A. 2024 The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. *Rev. Manag. Sci*. **18**, 1189–1220. (doi:10.1007/s11846-023-00696-z)

18. Mariani MM, Machado I, Magrelli V, Dwivedi YK. 2023 Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. *Technovation* **122**, 102623. (doi:10.1016/j.technovation.2022.102623)

19. Papadakis T, Christou IT, Ipektsidis C, Soldatos J, Amicone A. 2024 Explainable and transparent artificial intelligence for public policymaking. *Data Policy* **6**, e10. (doi:10.1017/dap.2024.3)

20. The Royal Society. 2023 *Science in the age of AI*. See https://royalsociety.org/news-resources/projects/science-in-the-age-of-ai/ (accessed 29 July 2024).

21. Valle-Cruz D, Criado JI, Sandoval-Almazán R, Ruvalcaba-Gomez EA. 2020 Assessing the public policy-cycle framework in the age of artificial intelligence: from agenda-setting to policy evaluation. *Gov. Inf. Q*. **37**, 101509. (doi:10.1016/j.giq.2020.101509)

22. Wang H *et al*. 2023 Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60. (doi:10.1038/s41586-023-06221-2)

23. Birhane A, Kasirzadeh A, Leslie D, Wachter S. 2023 Science in the age of large language models. *Nat. Rev. Phys*. **5**, 277–280. (doi:10.1038/s42254-023-00581-4)

24. Grossmann I, Feinberg M, Parker DC, Christakis NA, Tetlock PE, Cunningham WA. 2023 AI and the transformation of social science research. *Science* **380**, 1108–1109. (doi:10.1126/science.adi1778)

25. Bartneck C, Lütge C, Wagner A, Welsh S. 2021 *An introduction to ethics in robotics and AI*. Cham, Switzerland: Springer Nature. See https://library.oapen.org/handle/20.500.12657/41303 (accessed 29 July 2024).

26. European Commission. 2018 *A definition of Artificial Intelligence: main capabilities and scientific disciplines*. See https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines (accessed 30 July 2024).

27. Google. Google AI. 2024 *Google Responsible AI Practices*. See https://ai.google/responsibility/responsible-ai-practices/ (accessed 1 August 2024).

28. Köves A, Feher K, Vicsek L, Fischer M. 2024 Entangled AI: artificial intelligence that serves the future. *AI Soc*. (doi:10.1007/s00146-024-02037-4)

29. Arnold RD, Wade JP. 2015 A definition of systems thinking: a systems approach. *Procedia Comput. Sci*. **44**, 669–678. (doi:10.1016/j.procs.2015.03.050)

30. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. 2020 The Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med. Informatics Decis. Mak*. **20**, 310. (doi:10.1186/s12911-020-01332-6)

31. Bates J, Kennedy H, Medina Perea I, Oman S, Pinney L. 2024 Socially meaningful transparency in data-based systems: reflections and proposals from practice. *J. Doc*. **80**, 54–72. (doi:10.1108/jd-01-2023-0006)

32. Devlin J, Chang MW, Lee K, Toutanova K. 2019 BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc. of the 2019 conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (eds J Burstein, C Doran,

T Solorio), pp. 4171–4186, vol. 1. Minneapolis, MN: Association for Computational Linguistics. See https://aclanthology.org/N19-1423 (accessed August 2024).

33. Krizhevsky A, Sutskever I, Hinton GE. 2017 ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (accessed 16 August 2024).

34. Li J, Lim K, Yang H, Ren Z, Raghavan S, Chen PY, Buonassisi T, Wang X. 2020 AI applications through the whole life cycle of material discovery. *Matter* **3**, 393–432. (doi:10.1016/j.matt.2020.06.011)

35. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. 2023 Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.* **55**, 1–35. (doi:10.1145/3560815)

36. Sandbank J *et al*. 2022 Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. *NPJ Breast Cancer* **8**, 129. (doi:10.1038/s41523-022-00496-w)

37. Caldwell S *et al*. 2022 An *agile* new research framework for hybrid human-AI teaming: trust, transparency, and transferability. *ACM Trans. Interact. Intell. Syst.* **12**, 1–36. (doi:10.1145/3514257)

38. Pflanzer M, Traylor Z, Lyons JB, Dubljević V, Nam CS. 2023 Ethics in human–AI teaming: principles and perspectives. *AI Ethics* **3**, 917–935. (doi:10.1007/s43681-022-00214-z)

39. Kitchin R, Lauriault T. *Towards critical data studies: charting and unpacking data assemblages and their work*. Rochester, NY. See https://papers.ssrn.com/abstract=2474112 (accessed 17 August 2024).

40. Bates J, Lin YW, Goodale P. 2016 Data journeys: capturing the socio-material constitution of data objects and flows. *Big Data Soc.* **3**, 205395171665450. (doi:10.1177/2053951716654502)

41. Amoore L. 2020 *Cloud ethics: algorithms and the attributes of ourselves and others*. Durham, NC: Duke University Press.

42. Crawford K, Calo R. 2016 There is a blind spot in AI research. *Nature* **538**, 311–313. (doi:10.1038/538311a)

43. Thylstrup NB, Hansen KB, Flyverbom M, Amoore L. 2022 Politics of data reuse in machine learning systems: Theorizing reuse entanglements. *Big Data Soc.* **9**, 205395172211397. (doi:10.1177/20539517221139785)

44. Stigler SM. 2002 *Statistics on the table: the history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.

45. Fienberg SE. 1992 A brief history of statistics in three and one-half chapters: a review essay. *Stat. Sci.* (eds LJ Daston, G Gigerenzer, Z Swijtink, T Porter, L Daston, J Beatty), **7**, 208–225. (doi:10.1214/ss/1177011360)

46. Hahn GJ, Doganaksoy N. 2011 *The role of statistics in business and industry*. Hoboken, NJ: John Wiley & Sons.

47. Wild CJ, Pfannkuch M. 1999 Statistical thinking in empirical enquiry. *Int. Stat. Rev.* **67**, 223–248. (doi:10.1111/j.1751-5823.1999.tb00442.x)

48. Cox DR, Efron B. 2017 Statistical thinking for 21st century scientists. *Sci. Adv.* **3**, e1700768. (doi:10.1126/sciadv.1700768)

49. Medeiros L, Psaltis D, Lauer TR, Özel F. 2023 The image of the M87 black hole reconstructed with PRIMO. *Astrophys. J. Lett.* **947**, L7. (doi:10.3847/2041-8213/acc32d)

50. Jumper J *et al*. 2021 Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. (doi:10.1038/s41586-021-03819-2)

51. Idnay B, Dreisbach C, Weng C, Schnall R. 2021 A systematic review on natural language processing systems for eligibility prescreening in clinical research. *J. Am. Med. Informatics Assoc.* **29**, 197–206. (doi:10.1093/jamia/ocab228)

52. Gefen A, Saint-Raymond L, Venturini T. 2021 AI for digital humanities and computational social sciences. In *Reflections on artificial intelligence for humanity lecture notes in computer science* (eds B Braunschweig, M Ghallab), pp. 191–202. Cham, Switzerland: Springer International Publishing.

53. Cox A. 2021 The impact of AI, machine learning, automation and robotics on information professions. *CILIP*. See https://www.cilip.org.uk/page/researchreport (accessed 30 July 2024).

54. Jiang F *et al*. 2017 Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* **2**, 230–243. (doi:10.1136/svn-2017-000101)

55. Jorzik P, Klein SP, Kanbach DK, Kraus S. 2024 AI-driven business model innovation: a systematic review and research agenda. *J. Bus. Res.* **182**, 114764. (doi:10.1016/j.jbusres.2024.114764)

56. Bessant KC, MacPherson ED. 2002 Thoughts on the origins, concepts, and pedagogy of statistics as a 'separate discipline'. *Am. Stat.* **56**, 22–28. (doi:10.1198/000313002753631321)

57. Burnette DM. 2016 The renewal of competency-based education: a review of the literature. *J. Contin. High. Educ.* **64**, 84–93. (doi:10.1080/07377363.2016.1177704)

58. Alliance for Data Science Professionals. 2022 *The Alliance for Data Science Professionals Certification Guidance and Process: Advanced Data Science Professional*. See https://allianceofordatascienceprofessionals.co.uk/guidelines/latest (accessed 30 July 2024).

59. Kabalisa R, Altmann J *et al*. 2021 AI technologies and motives for AI adoption by countries and firms: a systematic literature review. In *Economics of grids, clouds, systems, and services lecture notes in computer science* (ed. K Tserpes), pp. 39–51. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-030-92916-9_4)

60. Rabiner L, Juang BH. 2008 Historical perspective of the field of ASR/NLU. In *Springer handbook of speech processing*, pp. 521–538. Heidelberg, Germany: Springer.

61. Savage N. 2020 How AI is improving cancer diagnostics. *Nature* **579**, S14–S16. (doi:10.1038/d41586-020-00847-2)

62. Minsky ML. 1962 Problems of formulation for artificial intelligence. In *Proc. of a Symp. on Mathematical Problems in Biology*, pp. 35–46. Providence, RI: American Mathematical Society. (doi:10.1090/psapm/014)

18

63. Binbasioglu M, Jarke M. 1987 Knowledge-based formulation of linear planning models. In *Expert systems and artificial intelligence in decision support systems* (eds HG Sol, C Takkenberg, PF De Vries Robbé), pp. 113–136. Dordrecht, The Netherlands: Springer. (doi:10.1007/978-94-009-3805-2_7)

64. Cramer EJ, Dennis Jr. JE, Frank PD, Lewis RM, Shubin GR. 1994 Problem formulation for multidisciplinary optimization. *SIAM J. Optim*. **4**, 754–776. (doi:10.1137/0804044)

65. Blagec K, Barbosa-Silva A, Ott S, Samwald M. 2022 A curated, ontology-based, large-scale knowledge graph of artificial intelligence tasks and benchmarks. *Sci. Data* **9**, 322. (doi:10.1038/s41597-022-01435-x)

66. Trewartha A *et al*. 2022 Quantifying the advantage of domain-specific pre-training on named entity recognition tasks in materials science. *Patterns* **3**, 100488. (doi:10.1016/j.patter.2022.100488)

67. Amironesei R, Denton E, Hanna A. 2021 Notes on problem formulation in machine learning. *IEEE Technol. Soc. Mag*. **40**, 80–83. (doi:10.1109/mts.2021.3104380)

68. Newman-Griffis D, Rauchberg JS, Alharbi R, Hickman L, Hochheiser H. 2023 Definition drives design: Disability models and mechanisms of bias in AI technologies. *First Monday* **28**. (doi:10.5210/fm.v28i1.12903)

69. Passi S, Barocas S. 2019 Problem formulation and fairness. In *Proc. of the Conf. on Fairness, Accountability, and Transparency (FAT* '19)*, Atlanta, GA, pp. 39–48. New York, NY: Association for Computing Machinery. (doi:10.1145/3287560.3287567)

70. Cheng L, Subrahmanian E, Westerberg AW. 2003 Design and planning under uncertainty: issues on problem formulation and solution. *Comput. Chem. Eng*. **27**, 781–801. (doi:10.1016/s0098-1354(02)00264-8)

71. Newman-Griffis D, Fosler-Lussier E. 2021 Automated coding of under-studied medical concept domains: linking physical activity reports to the international classification of functioning, disability, and health. *Front. Digit. Health* **3**, 620828. (doi:10.3389/fdgth.2021.620828)

72. Isagah T, Ben Dhaou SI. 2023 Problem formulation and use case identification of AI in government: results from the literature review. In *Proc. of the 24th Annual Int. Conf. on Digital Government Research*, pp. 434–439. New York, NY: Association for Computing Machinery. (doi:10.1145/3598469.3598518)

73. Ciarli T, Kenney M, Massini S, Piscitello L. 2021 Digital technologies, innovation, and skills: emerging trajectories and challenges. *Res. Policy* **50**, 104289. (doi:10.1016/j.respol.2021.104289)

74. Gaver WW. 1991 Technology affordances. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI '91)*, pp. 79–84. New York, NY: Association for Computing Machinery. (doi:10.1145/108844.108856)

75. Strauss LM, Klein AZ, Scornavacca E. 2024 Adopting emerging information technology: a new affordances process framework. *Int. J. Inf. Manag*. **76**, 102772. (doi:10.1016/j.ijinfomgt.2024.102772)

76. Chai J, Zeng H, Li A, Ngai EWT. 2021 Deep learning in computer vision: a critical review of emerging techniques and application scenarios. *Mach. Learn. Appl*. **6**, 100134. (doi:10.1016/j.mlwa.2021.100134)

77. Roman-Naranjo P, Parra-Perez AM, Lopez-Escamez JA. 2023 A systematic review on machine learning approaches in the diagnosis and prognosis of rare genetic diseases. *J. Biomed. Informatics* **143**, 104429. (doi:10.1016/j.jbi.2023.104429)

78. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. 2021 Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. **13**, 152. (doi:10.1186/s13073-021-00968-x)

79. Li Y, Wang S, Ding H, Chen H. 2023 Large language models in finance: a survey. In *Proc. of the 4th ACM Int. Conf. on AI in Finance*, Brooklyn, NY. New York, NY: Association for Computing Machinery. (doi:10.1145/3604237.3626869)

80. Li S, Wang G, Luo Y. 2022 Tone of language, financial disclosure, and earnings management: a textual analysis of form 20-F. *Financ. Innov*. **8**, 43. (doi:10.1186/s40854-022-00346-5)

81. Dalmer N, Newman-Griffis D, Ibrahimi M, Jia X, Allhutter D, Amelang K, Jarke J. 2024 Configuring data subjects. In *Dialogues in data power: shifting response-abilities in a datafied world* (eds J Jarke, J Bates), pp. 10–30. Bristol, UK: Bristol University Press. (doi:10.51952/9781529238327.ch001)

82. Gitelman L (ed). 2013 *'Raw data' is an oxymoron*. Cambridge, MA: The MIT Press. (doi:10.7551/mitpress/9302.003.0002)

83. Haraway DJ. 1991 *Simians, cyborgs, and women: the reinvention of nature*. See https://repository.library.georgetown.edu/handle/10822/545138 (accessed 22 August 2024).

84. Jordan MI, Mitchell TM. 2015 Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260. (doi:10.1126/science.aaa8415)

85. Chasalow K, Levy K. 2021 Representativeness in statistics, politics, and machine learning. In *Proc. of the 2021 ACM Conf. on Fairness, Accountability, and Transparency*, Virtual Event Canada, pp. 77–89. New York, NY: Association for Computing Machinery. (doi:10.1145/3442188.3445872)

86. Du B, Wang Z, Zhang L, Zhang L, Liu W, Shen J, Tao D. 2017 Exploring representativeness and informativeness for active learning. *IEEE Trans. Cybern*. **47**, 14–26. (doi:10.1109/TCYB.2015.2496974)

87. Agmon N, Ahituv N. 1987 Assessing data reliability in an information system. *J. Manag. Inf. Syst*. **4**, 34–44. (doi:10.1080/07421222.1987.11517792)

88. Gupta S, Gupta A. 2019 Dealing with noise problem in machine learning data-sets: a systematic review. *Procedia Comput. Sci*. **161**, 466–474. (doi:10.1016/j.procs.2019.11.146)

89. Eckert C, Isaksson O, Eckert C, Coeckelbergh M, Hagström MH. 2020 Data fairy in engineering land: the magic of data analysis as a sociotechnical process in engineering companies. *J. Mech. Des*. **142**, 121402. (doi:10.1115/1.4047813)

90. Eubanks V. 2018 *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York, NY: St. Martin's Publishing Group.

91. Ehsan U, Saha K, De Choudhury M, Riedl MO. 2023 Charting the sociotechnical gap in explainable AI: a framework to address the gap in XAI. *Proc. ACM Hum. Comput. Interact.* **7**, 1–32. (doi:10.1145/3579467)

92. Xu W, Dainoff MJ, Ge L, Gao Z. 2023 Transitioning to human interaction with AI systems: new challenges and opportunities for HCI professionals to enable human-centered AI. *Int. J. Human–Computer Interact.* **39**, 494–518. (doi:10.1080/10447318.2022.2041900)

93. Yu X, Xu S, Ashton M. 2023 Antecedents and outcomes of artificial intelligence adoption and application in the workplace: the socio-technical system theory perspective. *Inf. Technol. People* **36**, 454–474. (doi:10.1108/itp-04-2021-0254)

94. Mittelstadt B. 2019 Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **1**, 501–507. (doi:10.1038/s42256-019-0114-4)

95. Yoo D *et al*. 2024 A machine learning-driven virtual biopsy system for kidney transplant patients. *Nat. Commun.* **15**, 554. (doi:10.1038/s41467-023-44595-z)

96. UNOS. 2024 *How we match organs*. See https://unos.org/transplant/how-we-match-organs/ (accessed 23 August 2024).

97. Haggerty J, Roberge D, Pineault R, Larouche D, Touati N. 2007 Features of primary healthcare clinics associated with patients' utilization of emergency rooms: urban-rural differences. *Healthc. Policy* **3**, 72–85. (doi:10.12927/hcpol.2007.19394)

98. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. 2019 Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453. (doi:10.1126/science.aax2342)

99. Benton P. 2023 AI literacy: a primary good. In *Artificial intelligence research* (eds A Pillay, EJ Jembere, A Gerber), pp. 31–43. Cham, Switzerland: Springer Nature. (doi:10.1007/978-3-031-49002-6_3)

100. Domínguez Figaredo D, Stoyanovich J. 2023 Responsible AI literacy: a stakeholder-first approach. *Big Data Soc.* **10**, 20539517231219958. (doi:10.1177/20539517231219958)

101. Ng DTK, Leung JKL, Chu SKW, Qiao MS. 2021 Conceptualizing AI literacy: an exploratory review. *Comput. Educ.* **2**, 100041. (doi:10.1016/j.caeai.2021.100041)

102. Casal-Otero L, Catala A, Fernández-Morante C, Taboada M, Cebreiro B, Barro S. 2023 AI literacy in K-12: a systematic literature review. *Int. J. STEM Educ.* **10**, 29. (doi:10.1186/s40594-023-00418-7)

103. Kreinsen M, Schulz S. 2023 Towards the triad of digital literacy, data literacy and AI literacy in teacher education – a discussion in light of the accessibility of novel generative AI. *OSF*. See https://osf.io/xguzk (accessed 31 March 2024).

104. Southworth J, Migliaccio K, Glover J, Glover J, Reed D, McCarty C, Brendemuhl J, Thomas A. 2023 Developing a model for AI across the curriculum: transforming the higher education landscape via innovation in AI literacy. *Comput. Educ.* **4**, 100127. (doi:10.1016/j.caeai.2023.100127)

105. Long D, Magerko B. 2020 What is AI literacy? Competencies and design considerations. In *Proc. of the 2020 CHI Conf. on Human Factors in Computing Systems (CHI '20)*, Honolulu, HI. New York, NY: Association for Computing Machinery. (doi:10.1145/3313831.3376727)

106. Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, Hickey AJ, Clark AM. 2019 Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* **18**, 435–441. (doi:10.1038/s41563-019-0338-z)

107. Amodei D *et al*. 2016 Deep Speech 2: end-to-end speech recognition in English and Mandarin. In *Proc. of The 33rd Int. Conf. on Machine Learning. PMLR; 2016*, pp. 173–182. PMLR. https://proceedings.mlr.press/v48/amodei16.html.

108. Chib PS, Singh P. 2024 Recent advancements in end-to-end autonomous driving using deep learning: a survey. *IEEE Trans. Intell. Veh.* **9**, 103–118. (doi:10.1109/tiv.2023.3318070)

109. Glasmachers T. 2017 Limits of end-to-end learning. In *Proc. of the 9th Asian Conf. on Machine Learning. PMLR*, pp. 17–32. PMLR. https://proceedings.mlr.press/v77/glasmachers17a.html.

110. Paunov C, Planes-Satorra S, Ravelli G. 2019 *Review of national policy initiatives in support of digital and AI-driven innovation*. Paris, France: OECD. See https://www.oecd-ilibrary.org/content/paper/15491174-en (accessed 5 August 2024).

111. Essien A. 2023 AI-driven innovation: leveraging big data analytics for innovation. In *Innovation analytics*, pp. 119–137. London, UK: World Scientific (Europe). (doi:10.1142/9781786349989_0006)

112. Böckle M, Kouris I. 2023 Design thinking and AI: a new frontier for designing human-centered AI solutions. *Des. Manag. J.* **18**, 20–31. (doi:10.1111/dmj.12085)

113. Porter TM, Stone M. 2020 *The rise of statistical thinking, 1820–1900*. Princeton, NJ: Princeton University Press.