Technological Forecasting & Social Change 197 (2023) 122878

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Implementing and scaling artificial intelligence: A review, framework, and research agenda

Naomi Haefner[a,b], Vinit Parida[c,d,*], Oliver Gassmann[a,b], Joakim Wincent[a,e]

[a] Global Center for Entrepreneurship & Innovation, University of St. Gallen, Switzerland
[b] Institute of Technology Management, University of St. Gallen, Switzerland
[c] Entrepreneurship and Innovation, Luleå University of Technology, Sweden
[d] Department of Management/USN Business School, University of Vaasa/University of South-Eastern Norway, Finland/Norway
[e] Entrepreneurship and Innovation, Hanken School of Economics, Finland

A R T I C L E I N F O

Keywords:
Artificial intelligence
Machine learning
Review
Scaling
Innovation
Technology

A B S T R A C T

Artificial intelligence (AI) will have a substantial impact on firms in virtually all industries. Without guidance on how to implement and scale AI, companies will be outcompeted by the next generation of highly innovative and competitive companies that manage to incorporate AI into their operations. Research shows that competition is fierce and that there is a lack of frameworks to implement and scale AI successfully. This study begins to address this gap by providing a systematic review and analysis of different approaches by companies to using AI in their organizations. Based on these experiences, we identify key components of implementing and scaling AI in organizations and propose phases of implementing and scaling AI in firms.

1. Introduction

The release of ChatGPT and other generative artificial intelligence (AI) systems changed the rules of the game for businesses (Edelman and Abraham, 2023; OpenAI, 2022a). For several years now, experts have expected AI to have a far-reaching impact on virtually all industries (Berg et al., 2018; Chui et al., 2018). However, this new type of AI – generative AI – is supercharging these predictions (Chui et al., 2022). Generative AI includes large language models (e.g., LLaMA, see Meta AI, 2023; GPT-3, see OpenAI and Pilipiszyn, 2021; Bard, see Pichai, 2023), image-based systems (e.g., Midjourney, see Midjourney, 2022; DALL-E, see OpenAI, 2022b; Stable Diffusion, see Stability AI, 2022), and multimodal systems that combine different types of input (e.g., GPT-4, see OpenAI, 2023) as well as application-specific systems, such as AlphaFold for protein structure prediction (Hassabis, 2022). Anyone who has experimented with these systems can quickly see that they will enable more than just efficiency and efficacy improvements for businesses; they will create the basis for powerful new capabilities for firms (Chui et al., 2022). The largest tech firms, which are pushing the development of these foundation models (The Economist, 2022), are already integrating the technology into the core of their value propositions (Iansiti and Lakhani, 2020).

Many firms, however, clearly struggle to successfully implement AI. Recent surveys show that the vast majority of AI initiatives fail to take off (Browder et al., 2022; Ransbotham et al., 2020). Why is this? And how can firms gain traction with AI? To take full advantage of the potential benefits of AI technologies, firms must be able to successfully implement and ultimately scale AI in their organization. This requires going through a process of implementing and scaling AI. Firms must create the necessary prerequisites to successfully use AI technologies – that is to say, improve operational efficiency or build new value-creation capabilities based on AI technology. Without adapting to and adopting AI technologies, it will be difficult for many firms to remain competitive.

Given that implementing AI technologies is a necessity to stay competitive in the long run and that most firms continue to struggle with successfully deploying AI in their organizations, a detailed analysis of how firms can approach the implementation and scaling of AI in organizations is vital. Conversely, the implementation and scaling of AI in organizations has not been studied extensively thus far (Makarius et al., 2020). Moreover, companies lack guidance in managing these processes. Specifically, companies require frameworks to support the implementation and scaling of AI (Kanioura and Lucini, 2020). This study begins to address this gap by providing a systematic analysis of firms' explanations of their approaches to using AI in their organizations.

* Corresponding author at: Entrepreneurship and Innovation, Luleå University of Technology, Sweden.
E-mail addresses: naomi.haefner@unisg.ch (N. Haefner), vinit.parida@ltu.se (V. Parida), oliver.gassmann@unisg.ch (O. Gassmann), joakim.wincent@hanken.fi (J. Wincent).

https://doi.org/10.1016/j.techfore.2023.122878
Received 18 June 2022; Received in revised form 18 September 2023; Accepted 18 September 2023
Available online 5 October 2023
0040-1625/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Based on these experiences, we identify the key components of AI systems in organizations, and we pinpoint the phases that firms need to go through to implement and scale AI.

2. Theoretical background

2.1. Understanding AI adoption

AI, defined as “the science and engineering of making intelligent machines, especially intelligent computer programs” (McCarthy, 2007, p. 2), is gaining increasing relevance for firms. In particular, it is the relatively recent developments in machine learning, neural networks, and deep learning – approaches “concerned with the question of how to construct computer programs that automatically improve with experience” (Mitchell, 1997, p. xv) – that are driving substantial change for companies (Chui et al., 2018; Lakshmi and Bahli, 2020). Various authors have argued that the transformation toward the general use of AI in almost all company functions and areas is one of the most significant change drivers facing firms today (Acemoglu and Restrepo, 2020; Chalmers et al., 2021; Obschonka and Audretsch, 2020; Ransbotham et al., 2017; Zeba et al., 2021). The changes are expected to be far-reaching, touching virtually all aspects of firms’ business including decision making, manufacturing, marketing, supply chain management, logistics, recruiting, and more (Holmström, 2022; Makarius et al., 2020; Mishra et al., 2022; Murray et al., 2021; Pachidi et al., 2021; van den Broek et al., 2021; Wang and Su, 2021). Therefore, the potential from, and perhaps even the necessity of, adopting AI technologies are, for the most part, apparent to companies.

However, as with past phases when new technologies were introduced into organizations (Boothby et al., 2010), adopting AI is not as straightforward as many would hope. Very many companies struggle greatly to gain traction with AI (Browder et al., 2022; Ransbotham et al., 2020; Zolas et al., 2020). Part of the reason why firms struggle to implement AI is due to the fact that it “[differs] from other advanced technologies in [its] capacity to make determinations by [itself], as well as evolve [its] determinations over time once [it is] deployed in an organization” (Murray et al., 2021). This means that firms must be able to organize appropriately to ensure that the AI system can independently contribute to firm value creation as well as track and maintain these systems and their contributions to firm processes. Furthermore, other authors have pointed out that AI’s myopia (Balasubramanian et al., 2022) and inability to perceive interdependencies within the firm (Raisch and Krakowski, 2021) indicate that firms will likely encounter considerable difficulty in improving firm efficiency and expanding firm value creation using AI (Kemp, 2023).

Consequently, it is very important to better understand how AI implementation, and ultimately scaling, can be advanced in firms so that the considerable benefits inherent in the technology can be reaped. Research on technology adoption has pointed out that successful adoption hinges on the firm completing a transformation where it learns to both use the new technology and create the appropriate organizational setting. For instance, early studies on information technology (IT) adoption have shown that there is a “complementarity between computer investment and organizational investment” (Brynjolfsson et al., 2002, p. 138). More generally, the adoption of technology necessitates both changes in the firm’s technology use itself and in its organization (Boothby et al., 2010). AI technology is often discussed as an element in digital technologies more generally. The current wave of technological changes is significantly different from previous waves of technological development. Hanelt et al. (2021) note that current technologies are in and of themselves different from past technologies, such as IT. Current digital technologies are generative, malleable, and combinatorial (Kallinikos et al., 2013). Furthermore, digital technologies are reshaping firm boundaries and evoking more fundamental change even in firms’ business models (Hanelt et al., 2021). However, this literature has insufficiently examined the transformation process relating specifically to AI. Such an analysis – separate from other digital technologies, such as cloud computing – is essential because AI is effecting a fundamentally different advance for firms in “offload[ing] cognitive work from humans to computers” (Peretz-Andersson and Torkar, 2022, p. 2). Therefore, this paper examines the transformation of firms as they reach an initial readiness for AI and ultimately move forward to scale the use of the technology within their business. Companies that are known to be strong users of AI, such as Google and Uber, build their organizations around their AI systems (Johnson et al., 2022). Combining this practical knowledge on how AI-first firms operate with previous research on technology adoption, we seek to analyze the technological and organizational changes necessary for firms to embark on and ultimately complete the AI transformation successfully.

2.2. A socio-technical systems perspective on AI adoption

Socio-technical systems theory (STST) presents a very suitable framework for analyzing the implementation and scaling of AI in organizations because it has been used to study technology adoption more generally. It has been employed in different contexts including advanced manufacturing (Shani et al., 1992), renewable energy (Li et al., 2015; Yun and Lee, 2015), shipping (Geels, 2002), and mobile communications (Ansari and Garud, 2009; Shin et al., 2011). STST adopts a systems view of organizations – that is to say, it views any organization, or part of it, as consisting of a set of interacting sub-systems (Appelbaum, 1997; Geels, 2004). The theory is based on the socio-technical model developed by Leavitt (1965), whose original conceptualization proposed four interrelated and coordinated dimensions – namely, people, task, structure, and technologies – as the core components of organizational work systems. More recent work in the area of STST has tended to focus on a six-dimensional conceptualization consisting of people, goals, culture, infrastructure, technology, and processes (see e.g., Münch et al., 2022; Sony and Naik, 2020). Moreover, the external environment with the firm’s stakeholders and general market environment are often studied as part of STST (Münch et al., 2022).

STST is a suitable theoretical framework for our analysis of AI implementation and scaling in firms for several reasons. Our introduction to STST above shows that the theory presents a comprehensive framework to analyze emerging phenomena with interconnected technical and social dimensions. Critically, AI systems contain both technical and social aspects (Anthony et al., 2023; Glikson and Woolley, 2020; Lebovitz et al., 2022; Makarius et al., 2020). Therefore, STST is a highly appropriate framework to study AI implementation and scaling. Prior research in related areas such as servitization has shown that the joint study of social and technical components is beneficial in understanding the adoption and success of digital technologies (Münch et al., 2022; Sony and Naik, 2020).

Studies examining social and technical aspects of artificial intelligence specifically are, however, still very rare. The exceptions include Chowdhury et al. (2022) and Anthony (2021) who both examined socio-technical factors affecting individual employees’ collaboration with and use of AI. Xing et al. (2021) studied socio-technical barriers to the adoption of AI-based products by consumers. The only study to consider firm-level socio-technical aspects is Makarius et al. (2020), which proposes that the scope and novelty level of AI are key criteria in defining the approach to AI implementation. However, Makarius and colleagues note that “the fundamental issues related to structure and functioning in organizations in relation to AI systems remain underexplored” (Makarius et al., 2020, p. 271). Consequently, the goal of our research is to begin to address this precise gap in the literature by studying the process followed by companies in implementing and scaling AI in their organizations as well as the socio-technical components involved in this process.

3. Data and methods

We take an exploratory case study approach to examine the phases of implementing and scaling AI in a range of different companies. We rely on techniques for inductive theory building (Eisenhardt et al., 2016; Gehman et al., 2018). We base our approach to sample selection on that employed by Fisher et al. (2020), which uses podcasts featuring expert interviews as a data source. In our study, we used interviews conducted as part of the TWIML AI podcast (originally This Week in Machine Learning and AI), which is one of the top podcasts on artificial intelligence (Charrington, 2022). The podcast includes interviews with some of the top minds and ideas in machine learning and AI with the goal of elucidating the impact of AI on how businesses operate. The podcast guests include a wide range of machine learning and AI researchers, practitioners, and innovators. We analyzed podcast episodes that fulfilled the following criteria:

- The podcast featured a guest from industry practice rather than a researcher. This ensured that guests would be able to speak specifically to their experience with implementing and scaling AI applications in their respective firms.
- We selected a diverse set of firms that included both companies that are widely known to be among the most advanced in applying AI in their organizations and those that have relatively less experience. In doing so, we can capture a broader range of industry experience, with AI technologies applied in different settings.
- We included a wide variety of AI application areas from computer vision to sales forecasting, to content moderation. This approach allows us to derive more generally applicable approaches to using AI in organizations that go beyond the specifics related to particular AI technologies.

The firms analyzed in this study are all active in applying AI technologies. Importantly, they are in different phases of implementing and scaling AI (Table 1). Some interviewees have worked in more than one organization, allowing them to compare and contrast different approaches based on their varied personal experience. These cases represent real-life experience in applying AI technologies and managing the phases of implementing and scaling AI in different organizational contexts. Moreover, the interviewees provide some indication of what, in their experience, may constitute some of the best practices for implementing and scaling AI.

The study examines all interviews conducted with AI practitioners featured on the TWIML AI podcast between May 2021 and November 2022. The podcasts were analyzed regarding descriptions provided by practitioners on how they implemented AI technologies in their firms. We were intentionally broad in our initial analysis, including various types of information on using AI in firms without focusing on any specific aspect. We made notes on how experienced various practitioners and their companies were in applying AI technologies. This allowed us to determine how firms tend to progress through the phases of implementing and scaling AI.

Overall, we examined 22 interviews with machine learning and AI practitioners. Some interviewees providing insights into work both at their current firm and at well-known AI-first companies, such as Google, Uber, and Netflix, where they held prior positions. The practitioners held various positions including chief executive officer, chief data officer, head of AI, head of data science, and machine learning engineer. All interviewees had experience applying AI technologies in businesses and were, therefore, able to provide insights into the approaches chosen at their respective firms.

Our analysis of the interviews focused on recognizing key activities in AI implementation and scaling across the firms included in our study. We systematically uncovered themes in our complex data using thematic analysis (Braun and Clarke, 2006). This involved coding and categorizing phrases and themes mentioned by interviewees in the podcast episodes. This approach allowed us to identify various factors relevant to implementing and scaling AI in organizations. Subsequently, we examined the factors to identify underlying aggregate patterns. Finally, we mapped links between the aggregate patterns to derive a structured maturity model.

Table 1
Case description.

	Companies	Interviewee(s)	Industry
1	Toyota	Adrien Gaidon	Automotive
2	GSK	Kim Branson	Pharmaceuticals
3	Tecton/Uber	Mike del Balso Kevin Stumpf	Software/mobility
4	Cloudera	Sushil Thomas	Software
5	RTL	Daan Odijk	Media
6	23andMe	Subarna Sinha	Personal genomics and biotechnology
7	Intuit	Srivathsan Canchi	Financial software
8	LinkedIn	Ya Xu Parvez Ahammad	Professional network service
9	Overstock	Nishan Subedi	Internet retailer
10	Prosus	Paul van der Boor	Global investment group operating in social/gaming, classifieds, payments and fintech, edtech, food delivery, and ecommerce
11	ClearML/Google	Nir Bar-Lev	Software
12	AWS	Chris Fregly Antje Barth	Cloud computing
13	H&M	Errol Koolmeister	Fashion
14	Metaflow/Netflix	Ville Tuulos	Software/entertainment
15	Stack Overflow	Prashanth Chandrasekar	Software
16	Redfin	Akshat Kaul	Real estate
17	LEGO	Francesc Joan Riera	Toys
18	ADP	Jack Berkowitz	Software
19	Capital One	Ali Rodell	Finance
20	T-Mobile	Heather Nolis	Telecommunications
21	Preset/Airbnb	Maxime Beauchemin	Software/travel
22	Gojek	Willem Pienaar	Multi-service platform

4. Findings

This section presents the findings from our analysis of the interviews with machine learning and AI professionals. Section 4.1 describes the sub-dimensions of the socio-technical system for AI implementation and scaling. Then, Section 4.2 delineates how the socio-technical system is designed to accommodate firms initially implementing AI and ultimately scaling AI in their organizations.

4.1. Components of implementing and scaling AI

Implementing and scaling new technologies is a complex endeavor for firms (Fountaine et al., 2021). Recognizing which levers exist for management to successfully implement these technologies is, therefore, extremely important. Our analysis of the interviews in our study allowed us to determine the most important components outlined by the interviewees that increase the likelihood of successful implementation and scaling of AI in firms. We therefore employed an inductive approach to derive the key *technical* and *social* components involved in implementing and scaling AI in firms. In the following section, we briefly elaborate on the dimensions, elucidating the main sub-dimensions of the technical and social components.

4.1.1. Technical components

Managing the technology itself is clearly one of the most important aspects of implementing and scaling AI in organizations. Our analyses show that there are three levers within the technical components category. First, the *data pipeline* presents an intuitively important aspect. Many organizations struggle with collecting and maintaining the data necessary to fully implement and scale AI. Managing data is, therefore, an important aspect of enabling the implementation and scaling of AI. Second, the *technical infrastructure* refers to how and where AI systems are developed in organizations. There are different choices that firms can make in this area with implications for how well these AI systems can be scaled. Third, *AI models* indicate the types of algorithmic approach used by firms when developing their AI systems. These approaches can range from relatively simple traditional machine learning approaches to ensembles of highly sophisticated deep learning-based systems.

4.1.2. Social components

The other key dimension of successfully implementing and scaling AI in organizations involves setting the right social context. Here too, we see three important sub-dimensions of components. First, the *AI growth vision* refers to the overall goal set by the organization regarding using AI in the company. The vision sets the scope for choices taken within the other dimensions and sub-dimensions. Second, *AI capabilities* are key in successfully driving AI projects. Firms must bring together both technical and domain capabilities. Third, the *AI organizational structure* relates to the organizational design of the AI team and its responsibilities within the firm. Depending on the vision and the phase of implementing and scaling AI in the company, different approaches will be chosen.

These components describe the main levers available to firms attempting to implement and scale AI in their organizations. Our analysis shows that firms can develop different approaches to implementing and scaling AI along these technical and social dimensions.

4.2. Implementing and scaling AI in organizations

To successfully use AI, companies need to overcome certain hurdles and develop essential internal capabilities. We used the insights gleaned from machine learning and AI professionals at various companies to first determine the key socio-technical components required to implement and scale AI in firms. Next, we analyzed which capabilities these firms developed to overcome the hurdles identified and to successfully use AI in their respective firms.

This analysis of the interviews with machine learning and AI practitioners indicated that there were two main areas requiring development in firms to successfully use AI:

Develop the necessary technical infrastructure: Key technical developments are an important foundational driver for successful use of AI in companies. These requisite technical developments are closely related to the technical requirements of AI systems themselves. At a most basic level, training an AI system involves combining the following assets: collecting and managing *data pipeline* for training and validation, providing the necessary *technical infrastructure* to carry out the training and keep track of the deployed AI systems and, finally, hosting and maintaining a set of *AI models* used in the systems. Across these areas, firms have a variety of choices to make – for example, whether they make or buy certain stacks, or how to design the various elements modularly and flexibly.

Create the right social context: Having a suitable organizational setting within which to develop AI solutions is very important to successfully implement and scale AI in firms. First, organizations need to have an appropriate guiding *AI growth vision* that sets the scope and direction for AI implementation and scaling. Next, organizations must develop a range of *AI capabilities* that allow the firm to harness the potential of AI. This means building up both technical and domain capabilities. Often firms can supplement their internal capabilities with those offered by external partners to speed up the implementation and scaling of AI. Finally, firms need to set up the *organizational structure*. To successfully navigate AI implementation and scaling, firms need to combine the technical and business domain expertise needed to develop AI solutions and drive value creation for the business.

Our analyses of the interviews with machine learning and AI experts showed that there are various key activities in these two areas – technical and social components – supporting the implementation and scaling of AI in firms. We categorized the key activities according to different maturity levels to define the phases of AI implementation and scaling in firms (Table 2).

4.2.1. Level 1. Proving the concept

The first phase of implementing and scaling AI emphasizes the firm familiarizing itself with the potential of AI technologies and popularizing these within the organization. At this stage, firms can often rely on external support to implement initial AI solutions that clearly demonstrate the added value for the business.

Technical components: Our analyses show that, in this important first phase of implementing and scaling AI, creating a first *data pipeline* is highly important. This means starting to eliminate data silos and ensuring that the data infrastructure is ready for the next phases of implementing and scaling AI. Mike del Balso from Tecton/Uber noted: “A big challenge that a lot of these companies have is that they're still not kind of at ‘data maturity,’ so then building ‘ML maturity’ on top of that is a tricky spot to be in.” Without the necessary data pipeline maturity, therefore, the next phases will become significantly more difficult to reach.

Meanwhile, the approach to *technical infrastructure* can remain relatively simple in this first phase of implementing AI. Proof-of-concept AI applications can be developed on relatively modest infrastructure that is available to most firms. For example, Heather Nolis explained that T-Mobile started out with very rudimentary technical infrastructure: “We originally released our models in R. It was just R in a Docker container as an API. They ran pretty okay. We were doing two million returns a day.” Many firms also avail themselves of externally available services as offered, for example, by the large cloud providers. Daan Odijk from RTL described the advantages of using externally available services thus: “By taking these off-the-shelf models and then putting our own intelligence in the second level, our own labeling and where we have the data in that second level, that made a lot more sense for this learning problem as well.”

The latter has the advantage of allowing your AI teams to scale their systems more quickly in the future. Regarding *AI model* development, most machine learning and AI practitioners recommend starting with the easiest and simplest models possible. Errol Koolmeister describes the lessons learned in this regard at H&M:

> “When we started out and some of the consultancy early use cases and some of our early use cases, many of the people in the teams threw themselves directly into the latest research, wanted to do neural networks, wanted to get just a 0.1 uplift. But what we realized as well is that this isn't a Kaggle competition, it is not about optimizing the metrics and that, then you're done. It's about carrying it over into production into the infrastructure as well.”

Beyond relying on simpler models to achieve quicker gains, our analysis revealed that cloud providers often have services that allow firms to build their own applications based on existing models. Heather Nolis from T-Mobile recounted their own initial forays into model development:

> “I can speak about our speech-to-text here where we originally did roll out with vendor partners. We did a huge RFP (request for proposal). Every major speech-to-text provider in the world that exists, I have reviewed them. We launched our original proof of concept with AWS Transcribe. We did use a vendor. But immediately, once we had the audio data, we started looking at open source solutions and saying what can we do on our specific data.”

Table 2
Implementing and scaling AI with socio-technical components.

Components		Phases of Implementing and Scaling AI		
		Proving Concept	Productionizing	Platformizing
Technical	Data pipeline	Assemble data to run first use cases; enable data access; begin eliminating data silos	Organize your data repositories; enable access for variety of use cases; try to future-proof	Optimize for latency and throughput; democratize access
	Technical infrastructure	Simple on-premises setup; alternatively, rely on cloud providers	Central reference architecture; standardized and automated; tech agnostic; retain flexibility	Continually improve; eliminate pain points of existing applications
	AI models	Experiment with suitable approaches for your specific use cases; simple models over complicated ones; work with open-sourced models	Speed up experimentation; lower complexity; iterate continuously	Reuse capabilities (e.g., forecasting) in different contexts
Social	AI growth vision	Pursue valuable and feasible use cases; focus on efficiency	Customer value creation across value chain	Data-driven innovation
	AI capabilities	Start building technical capabilities; bring in external support for first use cases	Focus on speed; implement agile methods	Develop horizontal capabilities that work across business lines
	Organizational structure	Start building a central AI team; ensure close collaboration between technical and domain experts	Centrally organized AI team; vertical teams to support lines of business	Establish overarching, horizontal teams; increase (applied) research focus

This quotation touches on important learning regarding AI models, which is that firms should try to take advantage of the many open-source options available.

Social components: At the social level, the first phase of implementing AI requires the creation of an *AI growth vision* that focuses on identifying and successfully pursuing both valuable and feasible proofs of concept. To a certain extent, this allows firms to "pick the low hanging fruit" early on, which clearly demonstrates the potential benefits of AI technologies to a wide range of firm stakeholders, thereby improving company-wide buy-in. For example, Nishan Subedi from Overstock argued that "machine learning is, I think, best handled when there's at least clarity in terms of the objectives you want to achieve." At T-Mobile, meanwhile, Heather Nolis saw that it was important to "drive home a culture of small models for small problems. Build things specific for your use case to answer it exactly. Otherwise, you will get a deteriorated product."

In terms of *AI capabilities*, firms need to begin building their technical capabilities so that the first AI applications can be successfully implemented. Often, it can be a good tactic to look to external expertise – for example, from consulting companies or large cloud providers to help kick-start such early proofs of concept. Antje Barth from AWS explained the advantages of using Amazon SageMaker:

> "There have been a lot of additions to this managed service that is giving you basically the tools to build, to train, and to deploy models easily. At the same time, it's taking care of the infrastructure for you. It does the heavy lifting of managing individual instances. You can really focus on your tasks: to build models, to train the models, and to deploy them."

Importantly, the firm should begin to create the right *organizational structure*. In this phase, this means starting to build a central AI team. This will allow the firm to begin focusing on and developing its AI capabilities. Errol Koolmeister explains that: "the H&M approach was to do it centrally from the start basically to incubate the capability rather to spread it out." The organizational structure should enable good collaboration between the AI experts responsible for the technical implementation and the business domain experts who will be able to scope the various AI projects so that they can create real value for the organization.

4.2.2. Level 2. Productionizing

The second phase presents a relatively large leap forward in terms of the impact of implementing and scaling AI. At this stage, firms must set up the necessary technical and social components to allow AI systems to work *in production*, meaning that these systems are running and supporting a wide variety of business processes. The systems are being tracked and continuously updated while they provide a very large number of (real-time) inferences and predictions.

Technical components: In the second phase of implementing and scaling AI, it is important to have a *data pipeline* that is well organized so that it can be easily re-used in the future. Often, firms will optimize reusability by storing processed data that can then be reapplied in many use cases across the business as well as in possible new applications. Kim Branson from GSK noted that it is important to: "build data for future you so you can use it again. Collect those other data points at additional marginal cost that are really useful." Srivathsan Canchi from Intuit underlined this point:

> "As we were building ML models for servicing these different systems, we were discovering that we are building similar features because the data sets are highly intersecting. We have a lot of features that need to go across these systems and be shared between models; across turbo tax, quickbooks, and mint. To be effective at sharing such features, we needed a way to do that."

Moreover, the analysis revealed another important aspect regarding the data pipeline that firms faced during the "productionizing" phase. Specifically, the need to manage data access and ensure compliance increases during this phase. Subarna Sinha from 23andMe noted her company's complex requirements in this regard:

> "We need to make sure that we are satisfying research compliance. Then, we are also GDPR CCPA compliant. We have to have all of our training data... Even though we save our training data for a little bit, it expires. We have to have processes in place to destroy those buckets at a regular frequency and regenerate all of the training data as needed.

The exact requirements regarding data access and compliance vary by industry. Those companies operating in highly regulated industries, such as 23andMe, or those involving specific customers, such as children, must be able to address these issues very conscientiously in the "productionizing" phase.

The *technical infrastructure* is perhaps the most important pillar of the second phase of implementing and scaling AI. It must integrate all relevant components of AI systems from data management to experiment management, to orchestration and deployment management. In this phase, firms can develop both new AI systems and continuously run systems in production. A standardized infrastructure is the backbone enabling this dual exploration and exploitation in AI systems. Errol Koolmeister explains that H&M sees: "enormous productivity gain with designing an infrastructure that's reusable rather than building independent use cases at this scale." Firms need a deep understanding of the processes involved in developing AI systems, which allows them to standardize and automate many of the processes. Sushil Thomas from Cloudera explains that productionizing AI systems differs markedly from traditional web development:

"I think another challenge is just trying to understand what's different between that standard web app sort of development and product development versus like a ML/AI model going into production because there are large substantial differences that are really important to internalize and understand so that you can focus on different aspects of it as well."

In order to make broad use of AI across the organization, a solid technical infrastructure backbone is necessary. According to Akshat Kaul from Redfin, building sound technological infrastructure is key to using AI in different areas of the company:

"What we've been trying to do more recently is really develop that infrastructure, make it standardized, make it easy to use, really democratize machine learning within the company, and allow people teams across the business, across different domains to hire people who have that machine learning talent or to grow that talent and then use the platform that this team has built to tackle use cases in different domains."

Interestingly, machine learning and AI practitioners mention that it can be very helpful to be technologically agnostic and to retain a certain level of flexibility in the exact infrastructure choice – for example, regarding which cloud provider the firm chooses. This enables firms to be able to switch to better options as they become available, which is highly important in a field such as AI where many of the tools and services are still developing and improving on a regular basis. Ya Xu from Linkedin notes: "From [a] platform architecture standpoint, always think about how to build a platform in an extensible way." Jake Berkowitz points out why maintaining flexibility is necessary:

"The one thing you know about technology, somebody's going to come along with a better mousetrap. You need to be ready to re-architect and refactor. We think about it architecturally. We budget time for that as we go. It's the reality of the cloud."

Some companies such as GSK, have further, more specific criteria to build out their technical infrastructure. Kim Branson explains that: "the compute is really important for us. It's key to be unconstrained by compute." Such specific aspects will depend heavily on the use cases pursued by the particular company, but they should also be considered in developing the technical infrastructure.

In the "productionizing" phase of implementing and scaling AI, firms can speed up experimentation on new models as well as continuously improve existing *AI models*. Akshat Kaul from Redfin explains: "It's an ongoing effort. We are always working on improving the model that we have in production." Importantly, firms can often productionize their AI particularly well when they place the emphasis on simpler rather than more complex models. For instance, Errol Koolmeister from H&M recommends: "You don't go with the most complex technology or algorithm from the start because you don't know how that will scale." Combined with the ability to continuously improve models, the upshot is more reliable models overall. This aspect of the "productionizing" phase is especially relevant for a company such as ADP. Jack Berkowitz illustrates this point:

"The second thing is about reliability because we're messing with people's paychecks at the end of the day. That's what we do. That's the most personal data there is other than maybe healthcare data. If you want to see somebody get excited, make a mistake on their paycheck. So, we have to have a little bit of reliability in terms of what we're doing."

Reliability is evidently an important aspect to productionizing AI systems because companies increasingly need to make sure their AI systems provide accurate predictions and generate plausible output. Similarly, they need to be able to keep track of the consistency of their AI systems.

Social components: In this phase of implementing and scaling AI, the *AI growth vision* changes to target the application of AI across the value chain and throughout the entire organization. Kevin Stumpf describes the varied application areas for AI at Uber: "The use cases really varied from everything from supporting self-driving cars to dynamic pricing predictions to fraud detection, customer support, rider and order ETAs, restaurant recommendations." A key feature can be to prioritize the time to customer value creation, i.e., focusing on those applications and use cases where the organization can create real value for customers in a short time frame. Subarna Sinha describes how productionizing AI at 23andMe helped to change the growth vision for AI at the company:

"I think when you see something happen this quickly and you're able to say 'Oh I just want to train this and see what the results look like even if I'm not going to ship it to customers tomorrow.' It changes people... It's a delta, kind of a big shift in the way people think about it."

At this stage, it is important to focus on one key *AI capability*, which is speed. This involves ensuring that the organization can quickly experiment with new AI applications and test their performance in production. Adrien Gaidon explains how Toyota approaches this aspect of the "productionizing" phase:

"You want faster turnaround time and this kind of stuff. We want to create some kind of Toyota production system of deep learning. It's so that we can iterate really quickly from idea to model to validation and go back to the drawing board."

Machine learning and AI practitioners often work in agile ways, completing projects in sprints. This approach is even adopted by heavily research-oriented firms where the traditional pace is slower, and the development roadmap is oriented much more to the long term. For example, Kim Branson mentions that his company, GSK, "works in two-week sprints." Developing the capability to quickly iterate ideas and products is important to satisfy customer requirements as well.

A strong centrally positioned *organizational structure* can be a key component to transition to the second phase of implementing and scaling AI. Many companies focus on building a core AI team to support the various AI initiatives in the firm. Companies sometimes rely on additional vertical teams that closely support a particular line of business and help to solve problems specific to AI implementation in that line of business. Sushil Thomas from Cloudera explains:

"There are people who have a hub-and-spoke model where they have a central sort of center of excellence with the these are the guys who set up the best practices around the technologies they use and deployment practices and stuff like that. Then business units will have their own individual data scientists that they work with for their actual use cases."

As part of the organizational structure for the "productionizing" phase, it can be helpful to have key team members who evangelize for AI in the organization. These team members help to bring the AI growth vision to the entire organization. Ya Xu explains her approach at LinkedIn:

"The way that I also see that's worked really well is having this model that I like to call the champion model. Let's say for example you build a platform and you have to convince ten other teams to use it. Don't say hey all of you guys come and use my thing, but start with a couple of them who are already leaning in. They already showed interest, they already are actually excited about this thing."

Positioning champions for AI in the organizational structure can, therefore, be a very effective way to drive broad adoption of AI in the firm.

4.2.3. Level 3. Platformizing

The third phase of AI implementation allows firms to truly scale their AI systems. In "platformizing" AI in their organizations, businesses can take advantage of size, reusing capabilities in different contexts to fully ramp up AI in the organization.

Technical components: In this phase of scaling AI, firms rely very heavily on the previously established *data pipeline* and *technical infrastructure* approaches. To fully scale AI in the organization, both data management and infrastructure need to be optimized for latency and throughput. In this phase, AI systems must often make lightning-fast inferences and process vast amounts of data in real time. Indeed, during the "platformized" phase, companies often have to plan quite methodically for limitations in terms of inference speed and computing power. Ali Rodell points out how Capital One walks this tightrope:

> "A lot of people think that compute is unlimited because we work in the public cloud. Everybody reads compute is infinite in the cloud. One of the things we try and do is make sure that people understand that it is not. It is very scalable, but it is not unlimited."

The infrastructure should generally continue to improve in this phase to consider and eliminate any pain points encountered in running existing AI systems. To fully scale AI, it is helpful to democratize access to data and infrastructure so that AI solutions can be developed more quickly and by stakeholders with relatively less technical expertise. Kevin Stumpf from Tecton/Uber notes that: "The more data scientists you have, the more different use cases you have, the more sharing becomes a big part of it." Subarna Sinha shares the future potential she sees to expand AI at 23andMe:

> "The way we look at it, we feel we're really at kind of the beginning of what machine learning can do in terms of giving information on health. Right now a lot of our models primarily incorporate genetic information but there is room for incorporating information from wearables, information from lab values that you have, like your blood report or some measure of some other reports that you can have."

The most important technical component in the third phase of scaling AI is the way *AI models* are developed at this stage. Specifically, firms in this phase start to rely heavily on reusing capabilities. This means that they tend to develop "platforms" for related models. For instance, many firms find that forecasting is a problem that is highly relevant to many different business areas. At LinkedIn, such AI-based capabilities are reused strategically, as described by Ya Xu:

> "I know that you are quite familiar with our proML platform. You can actually build your modules on top of it. [...] A simple example is my team actually developed this model [for] explainability capability. They used it in their application. That went really well. They built it as an extensible module on the platform so other people who wanted to use it they can use it too."

In the "platformizing' phase of scaling AI, firms focus on building models that often open up the possibility of looking at a broader spectrum of related problems. Forecasting models, for instance, are germane to general time series problems, which can be applied in various company areas.

Social components: At the organizational level, the *AI growth vision* should now focus on the long-term strategy for AI to enable data-driven innovation in the organization. Targeted use cases should now consider reusability and applicability to multiple lines of business. Moreover, firms should consider where to strategically invest in capabilities that will likely become highly relevant for the business within the next few years. Parvez Ahammad from LinkedIn explains how the company thinks about this issue:

> "There are two investment pillars that I mentally use. One is how do we pick problems that, essentially, if we solve them very well, help multiple lines of businesses? Take, for example, something like doing a really good job on experimentation, helps multiple lines of businesses evaluate how their products are working. [It] helps them iteratively ship the products better. Something like forecasting, if you do a really good job, it also allows multiple businesses to have an ability to set their goals and actually measure how things are going and recognize when things are not going right. The utility across multiple lines of business is a key pillar for how we think about it. The second important pillar is what is the strategic scope or impact? On this pillar, it doesn't need to be actually a native part of the products today but something that we believe is going to be really important for us to invest in."

In terms of the *AI capabilities* required for this phase of scaling AI, here too, we see that firms are changing their emphasis from vertical capabilities specialized in business lines or functions to horizontal capabilities that can scale across business lines and functions. Sushil Thomas from Cloudera explains very succinctly that: "it's important to just uplevel org-wide what you can do with all of your data." Capabilities should be managed as a portfolio with an eye to balancing shorter-term capability development and more risky, longer-term capability development. According to Parvez Ahammad, LinkedIn wants "to be able to start new things or drop things that aren't working and be much more driven by market fit within the LinkedIn ecosystem." By strategically managing a portfolio of different AI projects, firms are able to exploit current capabilities while crafting space to create new ones.

To support the vision and capability development, firms in this phase of scaling AI often alter the *organizational structure*. Horizontal teams that work on overarching capabilities such as forecasting are required to fully scale AI platforms. These teams can have a more (applied) research focus to drive horizontal capabilities more effectively. Errol Koolmeister from H&M explains what the implications of the company's plans to scale AI are:

> "If we are 100 people today or 120-ish working on the AI use cases, we have around 10 use cases in production right now, if we're going to have all our core operational decisions amplified by AI by 2025, which is our tech leap that we're aiming towards, we're going to need thousands of people if we're scaling it vertically."

LinkedIn already operates its AI systems on a massive scale. Consequently, they have adapted their organizational structure to be able to effectively support the demands of these AI systems. Parvez Ahammad describes the teams at LinkedIn:

> "One of the things that is relatively new... couple of years ago we started these horizontal teams. There are at least a couple of them. One of them is called Data Science Applied Research, which is my team. There is a sister team that we have called Data Science Productivity and Experimentation."

LinkedIn has created a new organizational structure for AI at scale, which allows it to capitalize on the idea of reusable capabilities. The company achieved this by building teams on overarching topics such as applied research and experimentation. This setup allows LinkedIn to drive these topics across business units and application areas.

5. Discussion

5.1. Theoretical and management implications

The literature has repeatedly highlighted the potential of AI to dramatically increase firm performance, regardless of industry (Berg et al., 2018; Chui et al., 2018; Makarius et al., 2020). Some researchers have argued that the largest technology firms in particular have built a sustainable competitive advantage by shoring up their AI capabilities in digital platforms (Iansiti and Lakhani, 2020). At the same time, many other firms struggle to reap the rewards from AI (Browder et al., 2022; Ransbotham et al., 2020). This poses an interesting question of why some firms succeed while others fail to benefit from AI. Currently, pressure is increasing due to the rapid change induced by ChatGPT and other generative AI systems, which are substantially changing how businesses should be conducted in most industries (Edelman and Abraham, 2023). Even though experts have long predicted that AI will change the competitive arena and how to act in virtually all industries (Berg et al., 2018; Chui et al., 2018), this is now happening fast, and many firms are rapidly working on how best to implement and scale AI in their business. In reviewing how companies are coping with these challenges, we believe we are making several contributions to the research in this field.

Past work on technology adoption has shown that successful implementation centers on firms creating an appropriate socio-technical system within which to foster the new technology. This has been shown to be the case for IT adoption, for instance (Boothby et al., 2010; Brynjolfsson et al., 2002). Several studies indicate that AI technology contains socio-technical components that firms must manage in order to benefit from it (Anthony et al., 2023; Glikson and Woolley, 2020; Lebovitz et al., 2022; Makarius et al., 2020). Given that AI is generative, malleable, and combinatorial (Kallinikos et al., 2013), and that it can autonomously make predictions or generate outcomes and advance these over time (Murray et al., 2021), firms are faced with sizeable challenges in creating a suitable socio-technical system to benefit from the technology. However, research on the socio-technical components required to successfully exploit – that is to say, implement and scale – AI is still very limited, and there have been calls for further research in this area (Kanioura and Lucini, 2020; Makarius et al., 2020). The primary contribution of our study, therefore, is to begin to address this gap in the literature. Our exploratory qualitative approach allows us to gain a more fine-grained understanding of the socio-technical components of implementing and scaling AI. Specifically, we find that, on the technical side, firms should build out their data pipelines, technical infrastructure, and AI models. On the social side, firms should create an AI growth vision, build AI capabilities, and create an organizational structure to support the development of AI systems.

We further contribute to the literature by describing the phases of AI implementation and scaling. We find that the above socio-technical components play different roles during the process of moving from implementing to scaling AI. Firms generally begin with "proving the concept" of AI within the firm by launching their first viable use cases of the technology. Then, as firms plan to use AI more systematically, they advance to the "productionizing" phase. This phase is characterized by an increased standardization of processes to enable firms to handle more use cases. Finally, firms reach the "platformizing" phase, which allows them to run AI at scale. This means that they address use cases across the entire value chain, setting up the appropriate socio-technical components to *reuse capabilities* across AI applications.

Moreover, this study contributes to our understanding of how companies should act specifically when facing AI in their markets. We identify four basic advantages of implementing and scaling AI. First, implementing and scaling AI is considered a strategic necessity for most industrial sectors. As Nir Bar-Lev, the CEO of Clear ML, notes: "You probably want to have data scientists because if you want to be competitive, in virtually every industry today, you have to integrate AI into your business". This assertion is strongly supported by recent research on the applicability of AI, which indicates that AI can be employed in almost any sector (Chui et al., 2018). Consequently, at a most basic level, taking small steps to investigate how to implement and scale AI is a prerequisite for firms to remain competitive in the market. Firms that do not adopt AI risk being left behind.

Second, successfully implementing and, in particular, scaling AI are necessary for firms to fully realize the financial benefits that the technology can offer. Firms at the very early stages of AI implementation often struggle to do so (Ransbotham et al., 2020). These firms are seemingly stuck in proof-of-concept purgatory. The transition to productionized or platformized AI is difficult. Nir Bar-Lev explains how difficult it was for AI powerhouse, Google, to accomplish the transition only a few years ago when it worked to productionize an AI system that could help the company improve electrical consumption in its data centers by 40 % (Evans and Gao, 2016): "That got a lot of attention, but what wasn't known outside was that building that initial model took three weeks. Building a prototype to check it out, just validate it in one data center – not a working product, but a prototype – took three months. Rolling it out as a product? Over a year." Since there is still a considerable hurdle for firms to surmount in employing AI profitably, it is important to address this question (Browder et al., 2022). Bringing AI fully into the production environment allows firms to reap the financial benefits from these systems. This study furthers our understanding by identifying the key socio-technical components that firms scaling AI need to employ in their organizations.

Third, our study shows that productionizing and platformizing increase the speed, reliability, and explainability of AI systems. By standardizing and automating many of the core functions in AI systems, productionizing and platformizing improve the efficiency of AI. Jack Berkowitz, chief data officer at ADP, describes the main advantages his company sees in productionizing AI: "The first one is about pace. The world's busy, clients are demanding, and the world situations are changing all the time. The second thing is about reliability […] the third thing is really about clarity and explainability, whether it's to each other inside the development teams or whether it's to our end clients." By offering a systematic approach to AI development, productionizing creates the foundation for measurable and benchmarkable success. Such features are highly important because AI systems are often biased and it is, therefore, important to ensure that the systems run as intended and can be managed appropriately. This is perhaps especially true for generative AI systems that are prone to creating unreliable output. Consequently, firms should create the best possible socio-technical system to support their AI endeavors (Jackson, 2023).

Finally, our study indicates that platformizing AI enables entirely new opportunities for firms. This phase of scaling AI means that businesses can pursue novel data-driven innovation. The advanced techniques and reuse of capabilities create space for the firm to discover new prospects. Parvez Ahammad, head of data science applied research at LinkedIn, explains how working on one reusable capability has opened new paths for the company: "One of the nice side effects of getting the forecasting part right has been that we now are starting to look at a broader spectrum of time series problems. Forecasting is a sister problem to anomaly detection. Most people that come to you and ask for forecasting or anomaly detection also want a root cause analysis. There are a lot of statistical problems that are very adjacent. Initially, we were very focused on one, and when we got it right we feel like we have a p(0) answer. We are trying to slowly build something that's more holistic." Ultimately, the team thinks these developments will allow it to "change the culture of performance management [and make forecasting] a much more natural component of how people do matrix performance management." This example nicely illustrates how entirely new possibilities can be shaped by platformizing AI in organizations. In this phase, the potential is diverse, and firms will be able to create new data-driven initiatives supported by AI-based capabilities. This finding of our study is of particular interest for the AI literature on innovation and management because it indicates that AI may be a key way for firms to create capabilities going forward (Kemp, 2023). Our study is, therefore, among the first to show which socio-technical components are pertinent in creating AI-based capabilities and how firms can navigate the process of gaining competitive advantage using AI capabilities.

These benefits are findings uncovered from our analysis of the early stages. The full effects of scaling AI are presumably even more extensive. As noted, the implementation and scaling of AI in organizations is predicted to bring wide-ranging changes in essentially all sectors of industry and will substantially contribute to economic growth (Bughin et al., 2018; Chui et al., 2018). To fully take advantage of the potential benefits of AI, managers must pull the right socio-technical levers in their organizations. They need to develop the right technical and social components for AI.

5.2. Limitations and paths for future research

This study has some important limitations that, at the same time, point to certain interesting paths for future research. First and foremost, as a purely qualitative study, generalizability from our findings is somewhat limited. It would be a very worthwhile endeavor for future research to quantitatively examine the implementation and scaling of AI in organizations. In particular, it would be interesting for scholars to examine the extent to which employee knowledge and skills impact firms' ability to implement and scale AI because these factors are of considerable importance for firms attempting to benefit fully from modern technologies (Zheng and Hu, 2008). Relatedly, future work could consider studying how firms' plans to implement and scale AI interact with other firm strategies. For example, recent research has shown that some firm strategies are more conducive to good performance in turbulent market conditions (Beliaeva et al., 2020). Consequently, it may be fruitful to examine how other firm strategies support or impede technology adoption in firms.

Second, the literature on digital transformation, which is broader than AI transformation specifically, suggests that the external context can markedly influence the extent to which firms can and must transform (Hanelt et al., 2021). In line with this research, it may be helpful for future research to consider additional external conditions affecting firms' AI transformations. Future research may also benefit from analyzing the larger labor market implications of increased AI implementation and scaling. Past research has indicated that adoption of AI and robotics can have a significant impact on the labor market (Autor, 2015; Chen et al., 2022; Dixon et al., 2021). Since the implications of AI adoption are still ambiguous, with some studies suggesting that increased use of the technology will lead to lower employment and others arguing the opposite, it would be beneficial for future work to examine whether the AI transformation approach chosen by companies affects AI's impact on employees.

Finally, the current study did not limit itself in terms of industry, firm size, or geographical location. Future work could, therefore, examine how AI transformation is affected by the industry context, firm size, and sector location. Such analyses would allow the literature to establish more generalizable findings that hold across industry, firm size, and geography as well as to determine specific findings for smaller sets operating under special conditions.

6. Conclusion

The rapid advancement of AI presents significant challenges for companies, requiring them to adapt and navigate in order to integrate AI into their operations. To stay competitive, companies must proactively engage with AI technologies, strategically invest in talent and infrastructure, and establish a comprehensive AI framework to drive innovation and shape industry standards. This paper has argued that knowledge of AI implementation and scaling is rapidly needed and that it is time to review best practice. Our effort provides an initial framework and scientific model of AI implementation and AI scaling. Developing and implementing AI is known to be a difficult and challenging undertaking, but our accounts provide a list of benefits. Companies that succeed in implementing and scaling AI can ensure they remain competitive, drive value creation across the value chain, unlock efficient, reliable, and explainable AI solutions, and develop new AI-based capabilities. Success hinges on managing the implementation and scaling of AI by pulling the right socio-technical levers – developing a data pipeline, technical infrastructure, and AI models – while setting the right social context of laying out the AI growth vision, expanding AI capabilities, and establishing a suitable organizational structure. The framework we provide in this paper offers guidance on how to implement and scale AI to facilitate the successful use of AI in companies, leading the way to the next generation of highly innovative and competitive companies.

CRediT authorship contribution statement

Naomi HAEFNER: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Data Curation, Investigation, Methodology, Formal analysis, Visualization, Project administration, and Funding acquisition. **Vinit PARIDA:** Writing - Review & Editing, Investigation, Conceptualization, and Methodology. **Oliver GASSMAN**: Writing - Review & Editing, and Funding acquisition. **Joakim WINCENT:** Supervision, Conceptualization, and Editing.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors would like to express their sincere appreciation to The Research Council of Norway, the Swedish Energy Agency, and Wallander's och Tom Hedeliu's Stiftelse samt Tore Browaldhs Stiftelse for their funding support, which made this research possible.

References

Acemoglu, D., Restrepo, P., 2020. The wrong kind of AI? Artificial intelligence and the future of labour demand. Camb. J. Reg. Econ. Soc. 13 (1), 25–35. https://doi.org/10.1093/cjres/rsz022.

Ansari, S., Garud, R., 2009. Inter-generational transitions in socio-technical systems: the case of mobile communications. Res. Policy 38 (2), 382–392. https://doi.org/10.1016/j.respol.2008.11.009.

Anthony, C., 2021. When knowledge work and analytical technologies collide: the practices and consequences of black boxing algorithmic technologies. Adm. Sci. Q. 66 (4), 1173–1212. https://doi.org/10.1177/00018392211016755.

Anthony, C., Bechky, B.A., Fayard, A.-L., 2023. "Collaborating" with AI: taking a system view to explore the future of work. Organ. Sci. 34 (5), 1651–1996. https://doi.org/10.1287/orsc.2022.1651.

Appelbaum, S.H., 1997. Socio-technical systems theory: an intervention strategy for organizational development. Manag. Decis. 35 (6), 452–463. https://doi.org/10.1108/00251749710173823.

Autor, D.H., 2015. Why are there still so many jobs? The history and future of workplace automation. J. Econ. Perspect. 29 (3), 3–30. https://doi.org/10.1257/jep.29.3.3.

Balasubramanian, N., Ye, Y., Xu, M., 2022. Substituting human decision-making with machine learning: implications for organizational learning. Acad. Manag. Rev. 47 (3), 448–465. https://doi.org/10.5465/amr.2019.0470.

Beliaeva, T., Shirokova, G., Wales, W., Gafforova, E., 2020. Benefiting from economic crisis? Strategic orientation effects, trade-offs, and configurations with resource availability on SME performance. Int. Entrep. Manag. J. 16 (1), 165–194. https://doi.org/10.1007/s11365-018-0499-2.

Berg, A., Buffie, E.F., Zanna, L.-F., 2018. Should we fear the robot revolution? (The correct answer is yes). In: IMF Working Paper. https://www.imf.org/-/media/Files/Publications/WP/2018/wp18116.ashx.

Boothby, D., Dufour, A., Tang, J., 2010. Technology adoption, training and productivity performance. Res. Policy 39 (5), 650–661. https://doi.org/10.1016/j.respol.2010.02.011.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

van den Broek, E., Sergeeva, A., Huysman, M., 2021. When the machine meets the expert: an ethnography of developing ai for hiring. MIS Quarterly: Management Information Systems 45 (3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559.

Browder, R.E., Koch, H., Long, A., Hernandez, J.M., 2022. Learning to innovate with big data analytics in Interorganizational relationships. Academy of Management Discoveries 8 (1), 139–166. https://doi.org/10.5465/amd.2019.0048.

Brynjolfsson, E., Hitt, L.M., Yang, S., 2002. Intangible assets: computers and organizational capital. Brook. Pap. Econ. Act. 2002 (1), 137–198. https://doi.org/10.1353/eca.2002.0003.

Bughin, J., Seong, J., Manyika, J., Chui, M., Joshi, R., 2018. Notes from the AI frontier: Modeling the global economic impact of AI. In: McKinsey Global Institute, September. https://www.mckinsey.com/~/media/McKinsey/FeaturedInsights/ArtificialIntelligence/NotesfromthefrontierModelingtheimpactofAIontheworldeconomy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx%0A.

Chalmers, D., MacKenzie, N.G., Carter, S., 2021. Artificial intelligence and entrepreneurship: implications for venture creation in the fourth industrial revolution. Entrep. Theory Pract. 45 (5), 1028–1053. https://doi.org/10.1177/1042258720934581.

Charrington, S., 2022. About. TWIML AI Website. https://twimlai.com/about/.

Chen, N., Sun, D., Chen, J., 2022. Digital transformation, labour share, and industrial heterogeneity. J. Innov. Knowl. 7 (2), 100173. https://doi.org/10.1016/j.jik.2022.100173.

Chowdhury, S., Budhwar, P., Dey, P.K., Joel-Edgar, S., Abadie, A., 2022. AI-employee collaboration and business performance: integrating knowledge-based view, socio-technical systems and organisational socialisation framework. J. Bus. Res. 144, 31–49. https://doi.org/10.1016/j.jbusres.2022.01.069.
Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., Malhotra, S., 2018. Notes from the AI frontier: Insights from hundreds of use cases. In: McKinsey Global Institute. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning.
Chui, M., Roberts, R., Yee, L., 2022, December. Generative AI Is here: How Tools like ChatGPT Could Change your Business. McKinsey QuantumBlack. https://www.mckinsey.com/capabilities/quantumblack/our-insights/generative-ai-is-here-how-tools-like-chatgpt-could-change-your-business.
Dixon, J., Hong, B., Wu, L., 2021. The robot revolution: managerial and employment consequences for firms. Manag. Sci. 67 (9), 5586–5605. https://doi.org/10.1287/mnsc.2020.3812.
Edelman, D.C., Abraham, M., 2023. Generative AI Will Change your Business. Here's How To Adapt. Harvard Business Review Digital Articles. https://hbr.org/2023/04/generative-ai-will-change-your-business-heres-how-to-adapt.
Eisenhardt, K.M., Graebner, M.E., Sonenshein, S., 2016. Grand challenges and inductive methods: rigor without rigor mortis. Acad. Manag. J. 59 (4), 1113–1123. https://doi.org/10.5465/amj.2016.4004.
Evans, R., Gao, J., 2016. DeepMind AI Reduces Google Data Centre Cooling Bill by 40%. DeepMind Blog. https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40.
Fisher, G., Stevenson, R., Neubert, E., Burnell, D., Kuratko, D.F., 2020. Entrepreneurial hustle: navigating uncertainty and enrolling venture stakeholders through urgent and unorthodox action. J. Manag. Stud. 57 (5), 1002–1036. https://doi.org/10.1111/joms.12584.
Fountaine, T., McCarthy, B., Saleh, T., 2021. Getting AI to scale. Harv. Bus. Rev. 99 (3), 116–123. https://hbr.org/2021/05/getting-ai-to-scale.
Geels, F.W., 2002. Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. Res. Policy 31 (8–9), 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8.
Geels, F.W., 2004. From sectoral systems of innovation to socio-technical systems. Res. Policy 33 (6–7), 897–920. https://doi.org/10.1016/j.respol.2004.01.015.
Gehman, J., Glaser, V.L., Eisenhardt, K.M., Gioia, D., Langley, A., Corley, K.G., 2018. Finding theory–method fit: A comparison of three qualitative approaches to theory building. J. Manag. Inq. 27 (3), 284–300. https://doi.org/10.1177/1056492617706029.
Glikson, E., Woolley, A.W., 2020. Human trust in artificial intelligence: review of empirical research. Acad. Manag. Ann. 14 (2), 627–660. https://doi.org/10.5465/annals.2018.0057.
Hanelt, A., Bohnsack, R., Marz, D., Antunes Marante, C., 2021. A systematic review of the literature on digital transformation: insights and implications for strategy and organizational change. J. Manag. Stud. 58 (5), 1159–1197. https://doi.org/10.1111/joms.12639.
Hassabis, D., 2022. AlphaFold reveals the structure of the protein universe. Google DeepMind. July 28. https://www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-protein-universe.
Holmström, J., 2022. From AI to digital transformation: the AI readiness framework. Bus. Horiz. 65 (3), 329–339. https://doi.org/10.1016/j.bushor.2021.03.006.
Iansiti, M., Lakhani, K.R., 2020. Competing in the age of AI. Harv. Bus. Rev. 98 (1), 60–67. https://hbr.org/2020/01/competing-in-the-age-of-ai.
Jackson, B., 2023. Consider the Risks of Generative AI Before Adopting Game-Changing Tools. Forbes. March 3. https://www.forbes.com/sites/forbestechcouncil/2023/03/03/consider-the-risks-of-generative-ai-before-adopting-game-changing-tools/.
Johnson, P.C., Laurell, C., Ots, M., Sandström, C., 2022. Digital innovation and the effects of artificial intelligence on firms' research and development – automation or augmentation, exploration or exploitation? Technol. Forecast. Soc. Chang. 179, 121636. https://doi.org/10.1016/j.techfore.2022.121636.
Kallinikos, J., Aaltonen, A., Marton, A., 2013. The ambivalent ontology of digital artifacts. MIS Q. 37 (2), 357–370.
Kanioura, A., Lucini, F., 2020. A Radical Solution to Scale AI Technology. Harvard Business Review Digital Articles. https://hbr.org/2020/04/a-radical-solution-to-scale-ai-technology.
Kemp, A., 2023. Competitive advantages through artificial intelligence: toward a theory of situated AI. Acad. Manag. Rev. https://doi.org/10.5465/amr.2020.0205.
Lakshmi, V., Bahli, B., 2020. Understanding the robotization landscape transformation: A centering resonance analysis. J. Innov. Knowl. 5 (1), 59–67. https://doi.org/10.1016/j.jik.2019.01.005.
Leavitt, H.J., 1965. Applied organizational change in industry. In: March, J.G. (Ed.), Handbook of Organizations. Rand McNally, p. 1170.
Lebovitz, S., Lifshitz-Assaf, H., Levina, N., 2022. To engage or not to engage with AI for critical judgments: how professionals deal with opacity when using AI for medical diagnosis. Organ. Sci. 33 (1), 1–494. https://doi.org/10.1287/orsc.2021.1549.
Li, F.G.N., Trutnevyte, E., Strachan, N., 2015. A review of socio-technical energy transition (STET) models. Technol. Forecast. Soc. Chang. 100, 290–305. https://doi.org/10.1016/j.techfore.2015.07.017.
Makarius, E.E., Mukherjee, D., Fox, J.D., Fox, A.K., 2020. Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. J. Bus. Res. 120, 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045.
McCarthy, J., 2007. What is artificial intelligence? http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.
Meta AI, 2023, February 24. Introducing LLaMA: A Foundational, 65-Billion-Parameter Large Language Model. Meta AI Blog. https://ai.facebook.com/blog/large-language-model-llama-meta-ai/.
Midjourney. (2022, July 13). We're officially moving to open-beta! Join now at https://discord.gg/midjourney. **Please read our directions carefully** or check out our detailed how-to guides here: https://midjourney.gitbook.io/docs. Most importantly, have fun! [Tweet]. Twitter. https://twitter.com/midjourney/status/1547108864788553729.
Mishra, S., Ewing, M.T., Cooper, H.B., 2022. Artificial intelligence focus and firm performance. J. Acad. Mark. Sci. 50 (6), 1176–1197. https://doi.org/10.1007/s11747-022-00876-5.
Mitchell, T.M., 1997. Machine learning. In: Machine Learning, , 13th45. McGraw-Hill Science.
Münch, C., Marx, E., Benz, L., Hartmann, E., Matzner, M., 2022. Capabilities of digital servitization: evidence from the socio-technical systems theory. Technol. Forecast. Soc. Chang. 176 (February 2021) https://doi.org/10.1016/j.techfore.2021.121361.
Murray, A., Rhymer, J., Sirmon, D.G., 2021. Humans and technology: forms of conjoined agency in organizations. Acad. Manag. Rev. 46 (3), 552–571. https://doi.org/10.5465/amr.2019.0186.
Obschonka, M., Audretsch, D.B., 2020. Artificial intelligence and big data in entrepreneurship: A new era has begun. Small Bus. Econ. 55 (3), 529–539. https://doi.org/10.1007/s11187-019-00202-4.
OpenAI, 2022a. Introducing ChatGPT. OpenAI Blog. https://openai.com/blog/chatgpt.
OpenAI, 2022b. *DALL*-E Now Available in Beta. OpenAI Blog. July 20. https://openai.com/blog/dall-e-now-available-in-beta.
OpenAI, 2023. GPT-4. OpenAI. March 14. https://openai.com/research/gpt-4.
OpenAI, Pilipiszyn, A., 2021. GPT-3 Powers the Next Generation of Apps. OpenAI Blog. March 25. https://openai.com/blog/gpt-3-apps.
Pachidi, S., Berends, H., Faraj, S., Huysman, M., 2021. Make way for the algorithms: symbolic actions and change in a regime of knowing. Organ. Sci. 32 (1), 18–41. https://doi.org/10.1287/orsc.2020.1377.
Peretz-Andersson, E., Torkar, R., 2022. Empirical AI transformation research: A systematic mapping study and future agenda. E-Informatica Software Engineering Journal 16 (1), 220108. https://doi.org/10.37190/e-Inf220108.
Pichai, S., 2023. An Important Next Step on our AI Journey. Google: The Keyword. February 6. https://blog.google/technology/ai/bard-google-ai-search-updates/.
Raisch, S., Krakowski, S., 2021. Artificial intelligence and management: the automation–augmentation paradox. Acad. Manag. Rev. 46 (1), 192–210. https://doi.org/10.5465/AMR.2018.0072.
Ransbotham, S., Kiron, D., Gerbert, P., Reeves, M., 2017. Reshaping business with artificial intelligence. MIT Sloan Mangement Review 59 (1), 1–17. http://sloanreview.mit.edu/AI2017.
Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., LaFountain, B., 2020. Expanding AI's Impact With Organizational Learning. https://sloanreview.mit.edu/ai2020.
Shani, A. B.(Rami), Grant, R.M., Krishnan, R., Thompson, E., 1992. Advanced manufacturing systems and organizational choice: sociotechnical system approach. Calif. Manag. Rev. 34 (4), 91–111. https://doi.org/10.2307/41166705.
Shin, D.-H., Choo, H., Beom, K., 2011. Socio-technical dynamics in the development of next generation Mobile network: translation beyond 3G. Technol. Forecast. Soc. Chang. 78 (3), 514–525. https://doi.org/10.1016/j.techfore.2010.04.017.
Sony, M., Naik, S., 2020. Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. Technol. Soc. 61, 101248 https://doi.org/10.1016/j.techsoc.2020.101248.
Stability AI, 2022. Stable diffusion launch announcement. Stability AI Blog. August 10. https://stability.ai/blog/stable-diffusion-announcement.
The Economist. Huge "foundation models" are turbo-charging AI progress. June 11. https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress.
Wang, Y., Su, X., 2021. Driving factors of digital transformation for manufacturing enterprises: A multi-case study from China. Int. J. Technol. Manag. 87 (2/3/4), 229. https://doi.org/10.1504/IJTM.2021.120932.
Xing, F., Peng, G., Zhang, B., Li, S., Liang, X., 2021. Socio-technical barriers affecting large-scale deployment of AI-enabled wearable medical devices among the ageing population in China. Technol. Forecast. Soc. Chang. 166, 120609. https://doi.org/10.1016/j.techfore.2021.120609.
Yun, S., Lee, J., 2015. Advancing societal readiness toward renewable energy system adoption with a socio-technical perspective. Technol. Forecast. Soc. Chang. 95, 170–181. https://doi.org/10.1016/j.techfore.2015.01.016.
Zeba, G., Dabić, M., Čičak, M., Daim, T., Yalcin, H., 2021. Technology mining: artificial intelligence in manufacturing. Technol. Forecast. Soc. Chang. 171, 120971. https://doi.org/10.1016/j.techfore.2021.120971.
Zheng, C., Hu, M.-C., 2008. Challenge to ICT manpower planning under the economic restructuring: empirical evidence from MNCs in Singapore and Taiwan. Technol. Forecast. Soc. Chang. 75 (6), 834–853. https://doi.org/10.1016/j.techfore.2007.05.002.
Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D.N., Buffington, C., Goldschlag, N., Foster, L., Dinlersoz, E., 2020. Advanced technologies adoption and use by U.S. firms: evidence from the annual business survey. In: NBER Working Paper (No. 28290).

Naomi Haefner is an Assistant Professor of Technology Management at the University of St. Gallen, Switzerland. Her research interests include the impact of artificial intelligence on management and innovation in organizations. Her prior work has been published in Entrepreneurship Theory and Practice, Technological Forecasting and Social Change, and Small Business Economics.

Vinit Parida is a Professor in Entrepreneurship and Innovation at Luleå University of Technology, Sweden. His research interests include servitization, business models, open innovation, and organizational capabilities. He has published in journal articles such as, Academy of Management Journal, Strategic Management Journal, Journal of Management Studies, Long Range Planning, Industrial Marketing Management, Production and Operations Management, Journal of Cleaner Production, and others.

Oliver Gassman is Professor of Innovation Management at the University of St. Gallen, Switzerland. His-research interests include technology, business model innovation, open innovation, and innovation patterns. His-articles have been published in journals such as Journal of Management, Research Policy, Long Range Planning, Strategic Entrepreneurship Journal, Journal of Product Innovation Management, R&D Management, and others.

Joakim Wincent is Professor in Entrepreneurship and Management at Hanken School of Economics, Finland, and a Professor in Entrepreneurship and Innovation at University of St. Gallen. His current research interests include technology, artificial intelligence, management, and innovation in organizations. Previous articles by him have been published in journals such as the Academy of Management Review, Strategic Management Journal, Harvard Business Review, Journal of Management Studies, Journal of Business Venturing, Entrepreneurship Theory and Practice and others.