Bus Inf Syst Eng 66(2):221–231 (2024)
https://doi.org/10.1007/s12599-024-00851-0

CATCHWORD

Foundation Models

A New Paradigm for Artificial Intelligence

Johannes Schneider · Christian Meske · Pauline Kuss

Received: 19 May 2023 / Accepted: 3 December 2023 / Published online: 29 January 2024

Keywords Foundation models · Artificial intelligence · Generative AI · Emergent behavior · Prompting

1 Introduction

Recently, the domain of artificial intelligence (AI) has experienced a profound transformation with the emergence of foundation models as a new paradigm for developing AI systems (Bommasani et al. 2021). Foundation models constitute large-scale AI models that are pre-trained on vast amounts of general data and that can be adapted for downstream applications (e.g., by fine-tuning them through further training on application-specific data). Through this pre-train and adapt approach they expedite the development of innovative AI products and services and accelerate the accessibility of high-performance AI solutions in various industries (Teubner et al. 2023).

Foundation models show remarkable abilities to comprehend, generate, and adapt content across diverse domains, including creative generations (Chen et al. 2023), software debugging (Sobania et al. 2023), protein sequencing (Madani et al. 2023), or cross-modality outputs such as text-to-image creations (Ramesh et al. 2021). With scaling, foundation models are becoming increasingly good at performing tasks they were not explicitly trained for, thereby broadening the scope of applications achievable by a single model without the need for additional training data or fine-tuning (Brown et al. 2020). When needed, task-specific performance can be further enhanced through fine-tuning or effective prompt engineering techniques; both of which incur significantly lower costs in comparison to developing a new model from scratch (Liu et al. 2023; Niu et al. 2020).

The paradigm shift brought about by foundation models reshapes the design and deployment of AI applications. The advantages of large-scale foundation models including their emerging capabilities encourage convergence within the AI industry, leading to a growing number of AI applications being adjustments of only a few foundation models, owned by a few organizations and trained on a few datasets (Bommasani et al. 2021). Such homogenization promises great leverage to accelerate AI advancements across various domains. But it also raises concerns including monopolistic power structures, economic dependencies, or the potential dissemination of model vulnerabilities across a great number of downstream applications (Fishman and Hancox-Li 2022). As foundation models establish themselves as a cornerstone of state-of-the-art AI advancement, the dynamics of value creation and accumulation in the AI industry can be expected to change, and organizations might be forced to reconsider how they can differentiate their AI products and services in an age in which high-performance, multi-functional AI solutions are widely available. Lastly, because the design and control over AI systems is becoming dispersed across

Accepted after two revisions by Christine Legner.

J. Schneider (✉)
Department of Computer Science and Information Systems, University of Liechtenstein, Fuerst Franz Josef Str., 9490 Vaduz, Liechtenstein
e-mail: johannes.schneider@uni.li

C. Meske · P. Kuss
Faculties of Mechanical Engineering and Computer Science, Institute of Work Science, Ruhr University Bochum, Wasserstr. 221, 44799 Bochum, Germany
e-mail: christian.meske@rub.de

P. Kuss
e-mail: pauline.kuss@rub.de

an evolving ecosystem of actors, the shift from disparate models to a foundational approach challenges existing approaches of AI governance (Koniakou 2023; Schneider et al. 2023).

By redefining existing premises of AI development, management, and governance, the rise of foundation models will shape the trajectory of AI research, bringing forth important questions and opportunities for the field of Information Systems (IS) research and Business and Information Systems Engineering (BISE) (Dwivedi et al. 2023; Teubner et al. 2023). With this catchword article, we intend to contribute to the field's comprehension of foundation models and to outline a sociotechnical perspective (Sarker et al. 2019) on the intricate implications of this new paradigm for the construction and deployment of AI applications. To do so, we introduce the concept of foundation models and their defining features, followed by describing the implications of foundation models as a new paradigm of AI, and, finally, outlining opportunities for further IS research.

We begin Sect. 2 by defining the concept of foundation models and describing their emergence in the historical context of machine learning advancements to clarify their pivotal role in the current AI landscape. We then elaborate on the key features of foundation models – namely emergent capabilities, homogenization, and prompt sensitivity – and discuss the implications of these characteristics for AI development and deployment. In Sect. 3, we outline multiple avenues for future research, particularly focusing on opportunities relevant to the BISE community as a sociotechnical and construction-oriented discipline. In Sect. 4, we conclude the article with a summary and indicate its limitations.

2 Foundation Models as a New Paradigm for AI

Following Bommasani et al. (2021, p. 1), we define a *foundation model* as "any model that is trained on broad data that can be adapted to a wide range of downstream tasks". Through this combination of task-agnostic pre-training and subsequent fine-tuning, foundation models enable new approaches to building and deploying AI systems. Therefore, the rise of foundation models constitutes a paradigm shift in AI that promises unique potential and risks (Li et al. 2022), which is a key focus of this manuscript as elaborated in the following sections. Foundation models can be pre-trained on a specific modality (e.g., language, vision, robotics, reasoning) or show multi-modal capabilities (Reed et al. 2022). Current examples of foundational models include Open AI's GPT-Series (Brown et al. 2020), BERT, and CLIP. While entering numerous domains, the pre-training paradigm of foundation models takes shape most strongly in pushing benchmarks in the field of natural language processing (NLP) where applications such as ChatGPT, an application built on top of the GPT foundation model, can now generate texts that are increasingly indistinguishable from human writing (Bender et al. 2021).

Technically, foundation models are nothing new: they are based on long-existing deep neural networks and standard transfer learning. However, their size and large-scale training data result in newly emergent capabilities that can be transferred across applications and fine-tuned for the creation of numerous AI applications. Before introducing the defining features of foundation models and their respective implications for AI development and use, we describe the emergence of foundational models in the historical context of machine learning advancements.

2.1 History: From Expert Systems to Foundation Models

Foundation models can be seen as a "logical" step in the development of machine learning as shown in Fig. 1.

Before learning from data, decision-making by machines was done by expert systems which encoded explicit *rules extracted from experts* on how to turn an input into an output. Later, machine learning systems could operate "without explicitly being programmed", as coined by Samuel Jackson in 1959. At first, machine learning systems learned *decision rules* based on a set of criteria, i.e., features, still defined by experts. *(Simple) representation learning* aimed at automating the identification of features, which was marked by a breakthrough based on *deep learning* using artificial neural networks. Deep learning enabled machine learning systems to learn a hierarchy of features directly from the data, reducing the need for feature engineering (Janiesch et al. 2021). It demanded much larger datasets and models. Deep learning also allowed building models in a modular, flexible way by stacking various layers of artificial neurons on top of each other. This made it easy to enlarge models or to combine models trained on different modalities of data such as text and images. Consequently, deep learning was adopted in various areas of AI, including computer vision, speech, and natural language processing, and specific deep learning models were developed through different compositions of basic elements.

Initially, neural networks were mostly trained using supervised learning. Supervised learning relies on a labeled dataset in which each input is associated with an output label (Janiesch et al. 2021) commonly provided by a human. Model training was constrained by the availability of labeled training data. Addressing this challenge while also reducing training costs, the method of transfer learning

Model Size (#parameters)	None (Hand-designed rules)	Very few	Few	Large	Very large
Features	Hand-designed	Hand-designed	Learnt (Simple features)	Learnt (Hierarchy of features)	Learnt (Hierarchy of features)
Learning	None	Supervised	Supervised	Supervised	Self-supervised + in-context learning
Data	None	Very few labeled-data	Few labeled-data	Large labeled-data	Very large unlabeled data (+ small labeled data)
Adaptability	None	None	Little	Medium	Large

Fig. 1 History of machine learning

later allowed reusing the "knowledge" a neural network learned from one task (e.g., recognizing objects in images) for another task (e.g., recognizing behavior in videos) by keeping most of the model's architecture unchanged and retraining only parts of the network (Niu et al. 2020). Put differently, transfer learning enables pre-training a model on a surrogate task, and then adapting it via fine-tuning to a downstream task.

Transfer learning, together with scale, enabled the success of foundation models starting around 2017 (Bommasani et al. 2021). While transfer learning based on labeled datasets had been a practice before (e.g., Deng et al. 2009), the required pre-training was still limited by the cost of data annotation. This limitation was overcome by a wave of developments in self-supervised learning. In self-supervised learning the pre-training task is derived from task-agnostic data with self-generated labels. For example, language models such as the GPT-1 to GPT-4 are trained to predict the next word in a text, i.e., the label to predict is simply the next word in the text (OpenAI 2023a). While foundation models achieve state-of-the-art results that on some tasks outperform tailored models (Ziems et al. 2023), additional fine-tuning based on supervised learning (using a small-scale, task-specific dataset), reinforcement learning (using human feedback to generate output), or instruction tuning (based on a dataset described via instructions) often improves application-specific performance and model alignment with user intent (Ouyang et al. 2022). Fine-tuning of foundation models furthermore constitutes an important tool to increase models' alignment with ethical standards (Ouyang et al. 2022). For example, human feedback and additional datasets can be used to penalize toxic outputs or counterweight learned biases of a model.

Self-supervised learning hence revolutionized the utility of low-cost training data, which could now be abstracted through web crawling. This enabled the scaling of larger and more expressive pre-trained models that can be optimized in subsequent fine-tuning (Bender et al. 2021). Scaling was further supported by improvements in computer hardware and the development of the transformer model architecture (Vaswani et al. 2017). Unprecedented scalability and the resulting ability to efficiently handle large amounts of diverse data and to flexibly capture diverse information therefrom (model expressiveness) is what sets foundation models apart from prior deep learning models. This can be observed, for example, in models like GPT-4 (OpenAI 2023a) or Llama-2 which both perform extremely well on a broad set of tasks by increasing model and training data size compared to the original transformer model (Vaswani et al. 2017) and earlier models. Additional relevant properties of foundation models include multimodality (the processing of multimodal data such as images and text in one model), memory (storing and retrieving knowledge, possibly from a model external source), and compositionality (modularity of the model and generalizability) (Bommasani et al. 2021). Jointly, these properties result in the key features of foundation models such as emergent capabilities of in-context learning (Brown et al. 2020). In-context learning refers to models' ability to solve tasks without explicitly being trained on them, constituting a key feature of foundational models as detailed below.

2.2 Key Features of Foundation Models

To frame our consideration of foundation models from the perspective of IS research, we start with a description of their essential characteristics. Following previous research, we identify emergent capabilities and homogenization as key features of foundation models (Bommasani et al. 2021; Fishman and Hancox-Li 2022). We furthermore add prompt sensitivity, given its contingency on foundation model pre-training and its critical implications for AI development and deployment as detailed in Sect. 2.3.

Emergent Capabilities The first key feature of foundation models is the emergence of behavioral capabilities that were not explicitly constructed, nor expected, by human developers, which frequently is referred to as *in-context learning* (Min et al. 2022). Emergent capabilities are primarily investigated in the context of large-language models (e.g., Wei et al. 2022). During training, the large-scale model extracts a rich set of patterns and broad skills from the diverse training data, for example abstracting "an understanding" of the vocabulary and grammar of a language from a text corpus without being trained for a specific task (e.g., translation or question answering). During application, it can then flexibly employ respective modeling to perform various downstream tasks by the provision of a prompt, i.e., a description of the task in natural language or through a visual representation and, possibly, a few examples. For instance, GPT-3 and later versions show strong performance on mathematical tasks and analogical reasoning although they were not specifically trained for such (Brown et al. 2020; Webb et al. 2023). In contrast to GPT-3 (being a model of 175 billion parameters), the smaller GPT-2 (1.5 billion parameters) does not show comparable capabilities of in-context learning, demonstrating the contingency of emergent capabilities on model scale (Bommasani et al. 2021).

The feature of emergent capabilities is essential to models' high performance on complex tasks such as natural language understanding and generation. The proficiency of large-scale foundation models at "tasks defined on-the-fly" suggests that the relevance of fine-tuning for task-specific performance might decrease as models grow further in size (Brown et al. 2020, p. 9). This would decrease the cost and required effort to deploy foundation models for an even wider range of downstream tasks. Moreover, in-context learning allows to improve model performance by crafting adequate prompts without parameter updates (frequently referred to as *prompt engineering*).

The property of emergent capabilities also raises concerns. Emergence implies a substantial uncertainty over the capabilities, flaws, and limitations of foundational models, making it hard to understand, explain, and predict their behavior and potential failure modes (Bommasani et al. 2021). This is illustrated by the extent to which slightly altered prompts may cause considerably different outcomes; that is, small changes to the model input can cause a large change in its qualitative behavior. In this context, the complexity and relevance of prompt engineering can be expected to increase as new, unexpected model behavior emerges, a point we will return to below. Uncertainty over model behavior particularly demands caution where undesired behaviors of a foundation model could be passed on to adapted models downstream, implying that the key features of emerging properties and homogenization "interact in a potentially unsettling way" (Bommasani et al. 2021, p. 6).

Homogenization Foundation models give rise to unprecedented degrees of *homogenization*, referring to the consolidation of methodologies and models across AI applications and research communities. Homogenization is visible in three interrelated developments: new AI models are adjustments of (i) a few foundational models, (ii) trained on a few datasets, and/or (iii) by a few organizations (Bommasani et al. 2021). Homogenization is driven by the immense costs of training foundation models (e.g., computational and data aggregation costs), the monopoly of few companies on some large-scale proprietary datasets (e.g., social media platforms), and the self-perpetuating cycle of improved model performance, user engagement, and model-improving user feedback. These factors push toward a winner-takes-it-all dynamic that can be expected to result in the development of a few large-scale foundation models upon which a large share of future AI systems will be built.

Homogenization promises great leverage, as advances in the foundation model are automatically inherited by downstream AI systems. This leverage accelerates and decreases the cost of developing task-specific AI systems, now requiring only small-scale training or skillful prompting without any fine-tuning at all. Possibly, this makes AI available to new domains in which model training was previously unaffordable or impeded by the unavailability of rich datasets.

On the other hand, homogenization describes a centralization of field-leading AI research with a few companies, raising concerns related to power, dependencies, and the safeguarding of social and ethical interests within the economically driven private sector (Fishman and Hancox-Li 2022). Moreover, it risks algorithmic monoculture and implies that the same datasets – namely those underlying the training of the foundation models – are encoded within numerous AI applications (Kleinberg and Raghavan 2021). Intuitively, this points toward another major concern: homogenization leverages not only the benefits but also potential risks and flaws of the underlying model and datasets across the AI landscape (Bommasani et al. 2022). Put differently, any bias or undesirable behavior of a foundation model will, in the absence of preventative measures, likely be inherited by downstream applications.

Besides the centralization of AI advancements, homogenization also holds implications for the involved actors and their respective roles in developing AI applications (Hacker et al. 2023). Increasingly, AI applications will be the product of an emerging ecosystem comprised of foundation model providers, foundation model adapters and integrators, and end users, as illustrated in Fig. 2.

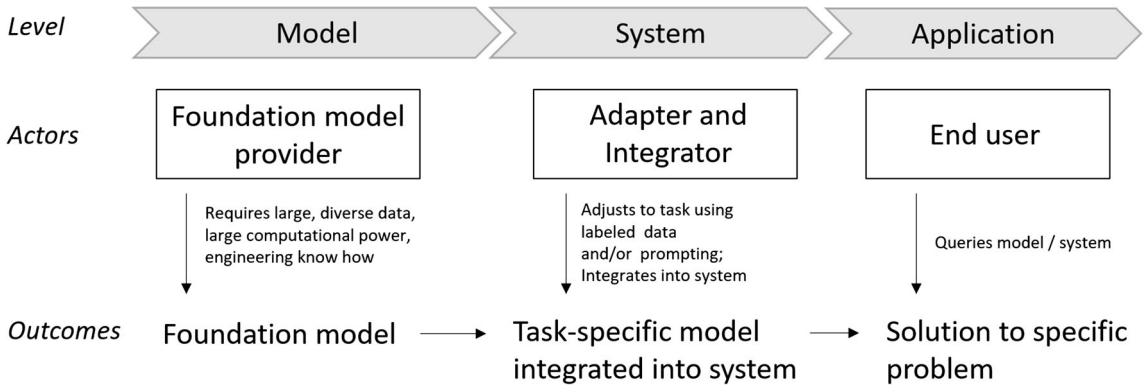

Fig. 2 AI development involving foundation models

Organizations operating as adapters and integrators fine-tune foundation models with labeled data or prompting creating task-specific models (Liu et al. 2023) and embed models into larger systems that are directly consumable by end users. Their respective choice of adaptation mechanism depends mostly on three factors: compute budget, data availability, and access to the foundation model. Adjusting the model itself to a specific task through transfer learning, using a (small) task-specific dataset, relies on the ability to modify model parameters directly or indirectly. Hence, it is contingent upon the provisions and allowances made by the foundation model providers, providing an example of power centralization through homogenization as noted above. Alternatively, adapters can fine-tune the behavior of a foundation model for a specific use case by means of prompting. Prompt-based fine-tuning does not depend on model access, but model providers can potentially restrict the permissibility of certain prompts. End users can similarly use prompting to impact model outputs. In the following section, we will consider the implications of this multilayered relevance of prompting in more detail.

Prompt Sensitivity Prompting refers to the use of natural language to guide the behavior of AI systems (Reynolds and McDonell 2021). Hence, the possibility of using prompts to instruct respective systems is contingent on their sophisticated capabilities to process natural language which, as detailed above, emerge from the foundation models pre-training on large-scale text corpora. However, the relevance of prompts for model behavior extends beyond pure language models and also encompasses multimodal foundation models such as text-to-image AI (Liu and Chilton 2022).

In the context of foundation models, prompting is highly relevant because the ability of the respective system to generate diverse and possibly objectionable output demands the possibility to specify and constrain desirable generations (Hacker et al. 2023). In this context, the exact wording of a prompt including instructions on how to tackle a problem (e.g., reasoning instructions, adoption of a role) and inclusion of exemplary input–output pairs (to leverage in-context learning) can have a significant impact on the behavior of respective systems. This makes prompt engineering, i.e., optimizing the most suitable prompt to realize desired outputs, an important tool to fine-tune, regulate, and query AI applications based on task-agnostic foundation models (Liu et al. 2023). For example, Kojima et al. (2022) showed that by adding "Let's think step by step" to the textual instruction of a prompt and therewith requiring serializing reasoning, the model performed significantly better on various benchmarks. This has been an unexpected phenomenon that, once again, is only visible with large-scale models (Chowdhery et al. 2023). Other techniques of prompt engineering similarly improve model performance, including task specification by example or the use of memetic proxies (Reynolds and McDonell 2021). The art of prompting receives increasing (scholarly) attention including the emergence of frameworks for systematic, reusable solutions to prompt engineering (e.g., White et al. 2023).

Referring to the above illustration of actors involved in the development and use of AI systems built on foundation models (Fig. 2), prompts can be used by each to guide model behavior: by the foundation model provider, by the foundation model adapter, and by the end user. Publications on prompt engineering techniques frequently foreground the prompts of end users querying downstream applications such as ChatGPT or DALL-E. However, natural language prompts can also be used by providers and adapters to address the foundation model directly. Besides fine-tuning, model providers and adapters may use prompting to constrain unwanted model behavior (Ouyang et al. 2022). Respective prompts intended to control system outcomes are also referred to as system prompts and receive particular attention in the context of jailbreaks – i.e., intentional attacks intended to circumvent system prompts through adversarial user prompts (Wei et al. 2023;

Zou et al. 2023). Recent research demonstrates that the generation of successful attack queries can be automated, indicating an issue of major concern with language-based foundation models that still requires solving (Zou et al. 2023). The feature of prompt sensitivity is thus essential with respect to the customization, control, and use of foundation models and their downstream applications. More work to better understand respective possibilities and limitations can be expected.

2.3 Implications for AI Development, Deployment, and Use

The characteristics of emergent capabilities, homogenization, and prompt sensitivity have implications for the development, deployment, and use of foundation models and their downstream applications. Most prominently, they drive the emergence of a new AI ecosystem and increase the speed and accessibility of AI advancements while challenging organizational and regulatory control.

With respect to the development of AI the rise of foundation models is realized by a new technology stack that fuels the emergence of an intricate AI ecosystem and a new distribution of influence over critical products and services. Relevant actors include the developers of foundation models and downstream applications, as well as hosting and hardware providers. For developers of foundation models, access to computational resources to train their large-scale models is essential, making processing technology a critical bottleneck in the development of new foundation models. The advancement of foundation model performance further hinges on large, high-quality training datasets. The rise of large-scale models is therefore accompanied by newly emergent markets for data generation and annotation, typically characterized by precarious labor conditions (Veselovsky et al. 2023). Due to homogenization, foundation model providers are incentivized to win market share by installing their models as the backbone of user-facing AI applications. OpenAI's recent reduction of model usage prices by 50% exemplifies this want to boost deployment. In addition, a shift from product to platform business is visible: OpenAI recently announced the launch of a platform, similar to the Apple App Store, to provide access to and the possibility to publish AI applications based on the OpenAI foundation models (OpenAI 2023c). The repositioning of foundation model providers as platform owners will likely accelerate homogenization processes and strengthen their influence on downstream AI applications. The emerging AI ecosystem also sees the rise of open-source models such as the Llama series by Meta offering an alternative to proprietary foundation models. Open-source models reduce engineering costs for application developers and allow them more insights into and control over model internals, while raising challenges similar to those for traditional software package reuse (Jiang et al. 2023). It remains to be seen how open-source alternatives will compete against or find integration in AI platforms like the one announced by OpenAI.

The reuse of foundation models will likely become a more prevalent paradigm for the development of AI applications. AI Developers' engagement with model design and training can thus be expected to decrease. Compared to classical paradigms of software engineering, new skills and organizational workflows are needed to realize AI solutions, including expertise in prompt engineering, knowledge of model reuse and customization, and the handling of ethical and safety risks in modular AI applications. Particularly given the foreshadowed emergence of platforms providing a multitude of fine-tuned models, the need to build a custom AI solution from scratch, declines. Instead, stacking and connecting modular AI applications can be expected to emerge as the leading approach for developing AI solutions. The expected trend toward cross-connecting AI applications to develop use-case-specific solutions requires revisiting existing frameworks for risk assessment and mitigation of (foundation model) AI developers: forecasting and reducing the risks of a given AI model typically involves delineating the boundaries of its reach and capabilities. However, developers of foundation models or downstream applications might not be able to conclusively foresee how their models will be integrated with other models, and how such integration could create new risks, including emergent capabilities, or subvert capability restrictions that were intentionally implemented as safety envelopes (Asatiani et al. 2021). Lastly, AI platforms backed by foundation models will enable developers to engage as a community and to directly market their applications to users, suggesting the emergence of new business models and blurring lines between AI developers and AI users.

With respect to the deployment of AI by and within organizations, foundation models revolutionize the availability of deployment-ready, low-cost, and high-performance AI applications. Consequentially, organizations will increasingly shift from custom-built AI systems toward deploying, and possibly customizing, pre-trained AI models. Deployment of AI will hence require new skill profiles, including expertise in choosing and integrating the right models for internal use cases or prompt engineering and fine-tuning skills for model customization. For the deployment of large-scale foundation models, organizations need to consider not only technical issues such as performance and reliability, but also sociotechnical aspects including necessary changes to existing workflows, novel protocols for human-AI collaboration, and cultivating a suitable safety culture.

As experimentation and implementation of AI use cases become easier, faster, and economically less risky, deployment cycles will accelerate. Moreover, bottom-up adoption of publicly accessible AI tools might jeopardize the centralized, managerial selection, evaluation, and orchestration of AI solutions deployed in organizational workflows. This decentralization and the black-box nature of foundation models intensify existing challenges of organizational AI deployment, including system interoperability, explainability of decision outcomes, and the governance of blurring accountability boundaries (Benbya et al. 2020; Minkkinen et al. 2023). The property of emergent capabilities in foundation models further complicates questions of accountability and the prevention of harmful outcomes, as unwanted, model behavior might arise unexpectedly (Bender et al. 2021; Bommasani et al. 2021). Consequentially, continuous monitoring and the implementation of fast-response safety protocols for worst-case scenarios become ever more critical in AI deployment, as organizational abilities to accurately foresee and prevent potential risks decrease. In the context of effective risk mitigation, some consider the homogenization of AI models and the centralization of computing capacities in the hands of a few commercial providers a regulatory opportunity: regulatory frameworks could leverage the central position of foundation model and infrastructure providers, for example through know-your-customer obligations, rather than posing a high regulatory burden on each deploying organization individually (Mökander et al. 2023).

Lastly, the carbon emissions and energy consumption associated with the training and deployment of large-scale foundation models are gaining attention (Bender et al. 2021). Organizations hence need to consider ways to incorporate respective environmental impact into sustainability strategies and corporate responsibility efforts.

With respect to the use of AI, new skills will be required to realize the economic potential of using the emerging class of AI. Examples include the skill of effective prompting to elicit accurate and relevant model responses. Organizational knowledge is needed on the individual and organizational factors influencing prompting skills (e.g., employees' comprehension of NLP or the specific model deployed; domain-specific expertise). Rapid fact-checking constitutes another important skill for efficient and responsible use, given the unresolved issue of hallucinations with AI applications based on large-scale foundation models (Ji et al. 2023). Besides awareness of the risk of hallucinations, users need a possibility, and possibly auxiliary tools, to cross-reference information with reliable sources. As a consequence, novel usage patterns of AI might emerge, ultimately changing how individuals fulfill their (work) tasks. The use of AI will likely also be affected by individuals' perception of personal risks. The importance of minimizing end-user risk to incentivize model adaptation is illustrated by OpenAI's copyright shield through which the company takes on legal responsibility if customers should be accused of copyright infringements because of their use of OpenAI's models (OpenAI 2023c). Users are faced with a decreasing explainability of AI solutions, and it remains to be seen how this affects their willingness to integrate respective systems for various use cases in work and private life. As noted above, as personalization and cross-integration of AI applications become easier, the line between users and developers will blur. Users no longer require sophisticated technical understanding and programming skills to build AI applications but can use natural language to customize solutions that best serve their needs.

3 Opportunities for IS Research

In this section, we outline future research directions for IS research, focusing on opportunities relevant to the BISE community as a sociotechnical and construction-oriented discipline. Some opportunities overlap with those generally described for deep learning and generative AI, including challenges of AI explainability (Meske et al. 2022), the organizational deployment of (generative) AI (Feuerriegel et al. 2023), and arising ethical concerns such as fairness (Feuerriegel et al. 2020).

We structure our overview according to the implications of the foundation models paradigm for AI development, deployment, and use as indicated above. We start with I. design and implementation of AI applications under the paradigm of foundation models, II. business models and value creation in the evolving AI ecosystem, III. AI management and governance, and IV. ecological and ethical dilemmas of foundation models.

3.1 Design and Implementation of AI Applications Under the Paradigm of Foundation Models

Re-using existing foundation models can be expected to substantially change organizations' approaches to designing AI products and services. Instead of training large models themselves, organizations will focus on model fine-tuning or prompt engineering to adapt foundation models to their specific use cases. Foundation model providers will relevantly influence AI application design through their usage policies and development guidelines, as illustrated by OpenAI (OpenAI 2023b). Future research is required to better understand the implications of this within and across organizations including new development processes, expertise, or resources required to leverage the potential

and mitigate the organizational hazards of externally developed foundation models. In this context, existing approaches to assessing and addressing ethical and safety concerns during design and engineering, including solutions focused on model explainability (Meske et al. 2022) or safety envelopment (Asatiani et al. 2021), require revision. The fragmentation of control between foundation model providers and downstream application developers is recognized in the proposed EU AI Act, which will be decisive for the allocation of regulatory obligations. More knowledge is needed regarding the relevant factors organizations ought to consider when deciding whether to use a foundation model or a custom model, selecting a specific foundation model, and deciding how to integrate it for their respective use cases. For the latter, additional knowledge is required on how to evaluate if model customization is necessary, and whether fine-tuning or prompt engineering is more suitable for the context of a particular organization and use case. Moreover, research should derive prescriptive design knowledge regarding the construction of viable downstream AI applications and investigate how organizations can establish a competitive advantage to position themselves in the evolving AI technology stack. Exemplary research questions include:

- How does the paradigm of foundation models challenge or reform established structures and processes for AI product and service development?
- What design principles can be established regarding the design of viable AI applications based on proprietary or open-source foundation models, and regarding model customization through either prompt engineering or fine-tuning?
- How does the paradigm of foundation models impact the comparable advantages of different organizational types (e.g., start-ups, corporates, SMEs) for constructing viable AI applications?

3.2 Business Models and Value Creation in the Evolving AI Ecosystem

The paradigm of foundation models implies unprecedented accessibility of high-performance AI models, accelerating AI development cycles, inviting low-cost experimentation, and enabling new applications such as generative AI (Chen et al. 2023) or life science innovations (Madani et al. 2023), The development of foundation models and their seamless integration into downstream applications necessitates a complex ecosystem of stakeholders. This spans from computational and storage hardware suppliers to foundation model developers, hosting providers, and an array of service providers, including prompt engineering managers. Future research should investigate the dynamics of this ecosystem, including inter-actor dependencies, drivers of value creation and accumulation, and the consequential emergence of novel business model innovations including platform solutions as foreshadowed by OpenAI (2023c). Moreover, a better understanding is needed of how organizations can realize a competitive advantage and differentiate their AI products and services in a homogenizing AI landscape. Existing works, e.g., on AI as a service (Lins et al. 2021), might provide a starting point. Exemplarily research questions include:

- Which opportunities for business model innovation emerge within the AI ecosystem, currently characterized by model, data and provider homogenization?
- What are the drivers of value creation in the evolving AI ecosystem? Where can we expect the accrual of value? Where can we expect the commodification of products and services? If a move from product to platform business model will be realized by foundation model providers like Open AI, how will this impact respective dynamics?
- How can organizations realize competitive advantages in the emerging ecosystem, including differentiation of their AI products or services despite homogenization?

3.3 AI Management and Governance

The integration of foundation models within business processes or AI products and services raises novel questions for managing and governing the organizational AI landscape. On a macro level, the allocation of responsibilities and liabilities among the diverse stakeholders remains to be formalized. The regulatory discourse on how to foster responsible behavior and ensure ethical, legally compliant AI systems is ongoing, illustrated by negotiations on the EU AI Act. Some aspects of concerns parallel those of the general public debate on AI, such as what an AI model should be allowed to do (autonomously) or how to balance freedom of speech and censorship in an attempt to realize ethical and legally compliant outputs. For example, should it be allowed for an AI to tell a joke related to gender or religion? With respect to agency, questions relate to what tasks an AI should be allowed to conduct autonomously: driving a car, providing medical advice, opening bank accounts and conducting online business?

Besides identifying the risks arising with pre-trained large-scale models, considerations involve an assessment of which risks can be addressed most effectively by whom, and how such mitigation could be executed. This includes preventing that model vulnerabilities are passed on from foundation models to downstream applications (Fishman and Hancox-Li 2022). Future research ought to investigate

respective best practices and derive strategic knowledge on how organizations can maneuver the balance between caution and safety without stifling innovation. Within organizations, established structures and processes of AI management and governance might require adaptation to the foundational model paradigm. This includes the management of bottom-up implementations of easily available AI solutions, solutions to privacy and copyright issues, and new governance approaches to assess and constrain the risks arising from a growing reliance on opaque black-box models or inaccessible training data. Existing data and AI governance frameworks might be leveraged (e.g., Schneider et al. 2023). In this context, the relevance of prompt engineering and related best practices to facilitate viable and safe AI systems require further investigation. Moreover, new benchmarks ought to be developed for assessing model performance including approaches to measure and control sociotechnical effects of foundation models within organizational structures and processes. Exemplarily research questions include:

- How can organizational AI management contribute to the realization of the value of foundation models with respect to internal business processes, and product or service development?
- How does organizational AI governance (have to) adapt to the increasing relevance of (opaque) foundation models in internal business processes, and product or service development?
- How can the responsibility to mitigate the risks of foundation models, including ethical, legal, and business concerns caused by the properties of emergent properties and homogenization, be distributed between different actors of the AI technology stack in a way that fosters safe and aligned AI applications?

3.4 Ecological and Ethical Dilemmas of Foundation Models

As foundation models become increasingly integral to various industries, ensuring their ethical deployment and long-term viability becomes paramount. Suitable frameworks and effective corporate governance mechanisms are required to guide decision-making processes related to the responsible development and deployment of foundation models and downstream applications. Future research should inquire into how organizations define and uphold ethical principles in this context, including the resolution of potential ethical dilemmas. Respective research can inform the establishment of robust corporate responsibility guidelines and policy recommendations to incentivize industry-wide adoption. Future research should also explore the broader economic and social implications of foundation models and investigate opportunities for corporate governance to influence respective effects, including the impact on workforce dynamics, market competition, or economic value distribution. Training and deploying foundation models come with significant ecological costs (OECD 2022). More research is needed to holistically capture the environmental footprint of foundation models and their associated infrastructure, including energy consumption, carbon emission, and resource utilization throughout the AI lifecycle. Avenues for sustainability improvements should be devised, including technical solutions (e.g., optimizing model architecture, reducing redundancy in training data, enhancing model compression techniques) and structural solutions (e.g., shared computing resources among industry players, federate learning networks, shared standards for model evaluation). Moreover, researchers could explore effective structures for incentivizing organizations to prioritize sustainability and stimulating industry-wide engagement in eco-friendly practices. Exemplarily research questions include:

- How can the training and deployment of (fewer) foundation models be realized in such a way that it contributes to improving the sustainability of the AI industry?
- How can corporate sustainability strategies integrate the environmental costs of recurrent inferences of foundation models integrated into internal business processes, or the organization's products and services?
- How do the key features of foundation models challenge existing (corporate) frameworks and criteria for responsible AI development or deployment? How can these frameworks be adapted to address the emerging challenges (e.g., lack of explainability of black-boxed foundation models)?

4 Conclusion

This article has elucidated the emergence of foundation models as a transformative paradigm for AI. Foundation models promise unprecedented opportunities to advance the performance and accessibility of AI applications across various sectors and represent a significant shift in how near-future AI systems will be developed, deployed, and used. We delineated key features of foundation models such as emergent capabilities, homogenization, and prompt sensitivity. These features redefine the stakeholders and dynamics of the AI ecosystem, including a pull toward a centralization of power and the rise of regulatory challenges through the diffusion of accountability and control. For organizations, foundation models provide a remarkable chance to revolutionize their operations, services, and

products. Avenues for future research include investigating how foundation models change organizations' approaches to designing AI applications, business models and value creation dynamics in the evolving AI ecosystem, hurdles and remedies concerning AI management and governance, and the formulation of organizational strategy and practices to harness the potential of foundation models responsibly and sustainably.

While we aimed to address a diverse array of aspects relevant to the BISE community in the context of foundation models, we recognize the limitation that there might be nuanced issues (e.g., potential differences in foundation models pre-trained on language vs. images or other modalities; prompt engineering; impact of model scaling on performance and key features) that were not exhaustively examined given the complexity of the topic. Moreover, due to the rapid speed of current technological advancements, we acknowledge a temporal constraint in this article's insights. However, we propose that an understanding of the fundamental characteristics of the foundation model paradigm, as elucidated within this article, establishes a critical groundwork for forthcoming discussions. Ultimately, the outlined avenues for future research offer an agenda for the BISE community that sets the stage for impactful contributions toward the viable, responsible, and sustainable realization of AI systems in the age of foundation models.

Funding Open access funding provided by University of Liechtenstein.

References

Asatiani A, Malo P, Nagbøl PR, Penttinen E, Rinta-Kahila T, Salovaara A (2021) Sociotechnical envelopment of artificial intelligence: an approach to organizational deployment of inscrutable artificial intelligence systems. J Assoc Inf Syst 22(2):325–252

Benbya H, Davenport TH, Pachidi S (2020) Artificial intelligence in organizations: current state and future opportunities. MIS Q Exec 19(4)

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: can language models be too big? In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pp 610–623

Bommasani R, Creel KA, Kumar A, Jurafsky D, Liang PS (2022) Picking on the same person: does algorithmic monoculture lead to outcome homogenization? Adv Neur Inf Proc Syst 35:3663–3678

Bommasani R, Hudson DA, Adeli E et al (2021) On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258

Brown T, Mann B, Ryder N et al (2020) Language models are few-shot learners. Proc Adv Neural Inf Process Syst 33:1877–1901

Chen L, Sun L, Han J (2023) A comparison study of human and machine generated creativity. J Comput Inf Sci Eng 23(5):051012

Chowdhery A, Narang S, Devlin J et al (2023) Palm: Scaling language modeling with pathways. J Mach Learn Res 24(240):1–113

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition, pp 248–255

Dwivedi YK, Kshetri N, Hughes L et al (2023) "So what if chatgpt wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manag 71:102642

Feuerriegel S, Dolata M, Schwabe G (2020) Fair AI. Bus Inf Syst Eng 62(4):379–384

Feuerriegel S, Hartmann J, Janiesch C, Zschech P (2023) Generative AI. Bus Inf Syst Eng. https://doi.org/10.2139/ssrn.4443189

Fishman N, Hancox-Li L (2022) Should attention be all we need? The epistemic and ethical implications of unification in machine learning. In: Proceedings of the ACM conference on fairness, accountability, and transparency, pp 1516–1527

Hacker P, Engel A, Mauer M (2023) Regulating ChatGPT and other large generative AI models. In: Proceedings of conference on fairness, accountability, and transparency, pp 1112–1123

Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Mark 31(3):685–695

Ji Z, Lee N, Frieske R et al (2023) Survey of hallucination in natural language generation. ACM Comput Surv 55(12):1–38

Jiang W, Synovic N, Hyatt M et al (2023) An empirical study of pre-trained model reuse in the hugging face deep learning model registry. arXiv preprint arXiv:2303.02552

Kleinberg J, Raghavan M (2021) Algorithmic monoculture and social welfare. Proc Natl Acad Sci 118(22):e2018340118

Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y (2022) Large language models are zero-shot reasoners. Adv Neural Inf Proc Syst 35:22199–22213

Koniakou V (2023) From the "rush to ethics" to the "race for governance" in artificial intelligence. Inf Syst Front 25(1):71–102

Li X, Tian Y, Ye P, Duan H, Wang FY (2022) A novel scenarios engineering methodology for foundation models in metaverse. IEEE Trans Syst Man Cybern Syst 53(4):2148–2159

Lins S, Pandl KD, Teigeler H, Thiebes S, Bayer C, Sunyaev A (2021) Artificial intelligence as a service. Bus Inf Syst Eng 63(4):441–456

Liu V, Chilton LB (2022) Design guidelines for prompt engineering text-to-image generative models. In: Proceedings of the CHI conference on human factors in computing systems. https://doi.org/10.1145/3491102.3501825

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(9):1–35

Madani A, Krause B, Greene ER et al (2023) Large language models generate functional protein sequences across diverse families. Nat Biotechnol 41:1099–1106

Meske C, Bunde E, Schneider J, Gersch M (2022) Explainable artificial intelligence: objectives, stakeholders, and future research opportunities. Inf Syst Manag 39(1):53–63

Min S, Lyu X, Holtzman A, Artetxe M, Lewis M, Hajishirzi H, Zettlemoyer L (2022) Rethinking the role of demonstrations: What makes in-context learning work? In: Proceedings of the conference on empirical methods in natural language processing, pp 11048–11064

Minkkinen M, Zimmer MP, Mäntymäki M (2023) Co-shaping an ecosystem for responsible AI: five types of expectation work in response to a technological frame. Inf Syst Front 25(1):103–121

Mökander J, Schuett J, Kirk HR, Floridi L (2023) Auditing large language models: a three-layered approach. AI and Ethics. https://doi.org/10.1007/s43681-023-00289-2

Niu S, Liu Y, Wang J, Song H (2020) A decade survey of transfer learning (2010–2020). IEEE Trans Artif Intell 1(2):151–166

OECD (2022) Measuring the environmental impacts of artificial intelligence compute and applications: the AI footprint. OECD Digital Economy Papers, No. 341, OECD Publishing, Paris. https://doi.org/10.1787/7babf571-en

OpenAI (2023a) GPT-4 technical report. https://arxiv.org/abs/2303.08774

OpenAI (2023b) Usage policies. https://openai.com/policies/usage-policies. Accessed 28 Oct 2023

OpenAI (2023c) Introducing GPTs. https://openai.com/blog/introducing-gpts. Accessed 8 Nov 2023

Ouyang L, Wu J, Jiang X et al (2022) Training language models to follow instructions with human feedback. Adv Neural Inf Proc Syst 35:27730–27744

Ramesh A, Pavlov M, Goh G et al (2021) Zero-shot text-to-image generation. In: International conference on machine learning, pp 8821–8831

Reed S, Zolna K, Parisotto E et al (2022) A generalist agent. arXiv preprint arXiv:2205.06175

Reynolds L, McDonell K (2021) Prompt programming for large language models: beyond the few-shot paradigm. In: Extended abstracts of the CHI conference on human factors in computing systems, pp 1–7

Sarker S, Chatterjee S, Xiao X, Elbanna A (2019) The sociotechnical axis of cohesion for the is discipline: its historical legacy and its continued relevance. MIS Q 43(3):695–720

Schneider J, Abraham R, Meske C, Vom Brocke J (2023) Artificial intelligence governance for businesses. Inf Syst Manag 40(3):229–249

Sobania D, Briesch M, Hanna C, Petke J (2023) An analysis of the automatic bug fixing performance of ChatGPT. arXiv preprint arXiv:2301.08653

Teubner T, Flath CM, Weinhardt C, van der Aalst W, Hinz O (2023) Welcome to the era of ChatGPT et al. The prospects of large language models. Bus Inf Syst Eng 65(2):95–101

Vaswani A, Shazeer N, Parmar N et al (2017) Attention is all you need. Proc Adv Neural Inf Process Syst 30:5999–6009

Veselovsky V, Ribeiro MH, West R (2023) Artificial artificial artificial intelligence: crowd workers widely use large language models for text production tasks. arXiv preprint arXiv:2306.07899

Webb T, Holyoak KJ, Lu H (2023) Emergent analogical reasoning in large language models. Nat Hum Behav 7:1526–1541

Wei J, Tay Y, Bommasani R et al (2022) Emergent abilities of large language models. Trans Mach Learn Res. https://doi.org/10.48550/arXiv.2206.07682

Wei A, Haghtalab N, Steinhardt J (2023) Jailbroken: how does LLM safety training fail? arXiv preprint arXiv:2307.02483

White J, Fu Q, Hays S et al (2023) A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382

Ziems C, Held W, Shaikh O, Chen J, Zhang Z, Yang D (2023) Can large language models transform computational social science? arXiv preprint arXiv:2305.03514

Zou A, Wang Z, Kolter JZ, Fredrikson M (2023) Universal and transferable adversarial attacks on aligned language models. arXiv preprint arXiv:2307.15043