How Text-to-Image Generative AI Is Transforming Mediated Action

Henriikka Vartiainen and Matti Tedre, *University of Eastern Finland, FI-80101, Joensuu, Finland*

This article examines the intricate relationship between humans and text-to-image generative models (generative artificial intelligence/genAI) in the realm of art. The article frames that relationship in the theory of mediated action—a well-established theory that conceptualizes how tools shape human thoughts and actions. The article describes genAI systems as learning, cocreating, and communicating, multimodally capable hybrid systems that distill and rely on the wisdom and creativity of massive crowds of people and can sometimes surpass them. Those systems elude the theoretical description of the role of tools and locus of control in mediated action. The article asks how well the theory can accommodate both the transformative potential of genAI tools in creative fields and art, and the ethics of the emergent social dynamics it generates. The article concludes by discussing the fundamental changes and broader implications that genAI brings to the realm of mediated action and, ultimately, to the very fabric of our daily lives.

Generative AI is at the forefront of technological advancements in artificial intelligence (AI). The term refers to systems that create new, original content, such as images, text, video, or music, based on patterns they have learned from existing examples. Large generative AI models that have been trained with massive amounts of data—often called "foundation models"—are especially attractive due to their adaptability to a wide range of more specialized downstream tasks and the accompanying potential for transformative effects in industries and domains ranging from art and music to literature, design, healthcare, education, and much more.[1]

Within this development, generative AI has attracted visions and imaginaries of how it provides new means to foster imagination, self-expression, and networked cocreativity. Simultaneously, it has also raised concerns and discussions about the extent to which generative AI tools are used to automate creative work and disrupt job markets in the creative fields.[1] Similarly, there have been concerns about the consequences of using generative AI in the education sector, particularly because students may use AI tools to complete their assignments without any effort of their own, but also due to risks associated with the homogenization of education, centralization of decision-making, and diminished diversity.

Even amid growing concerns, uncertainty, and resistance surrounding generative AI, the proliferation of generative AI tools makes it increasingly necessary to consider how these tools reshape people's cultural practices and the ways individuals and communities engage with the world. It becomes important not only to conceptualize and theorize the nature of these changes, but also to remain mindful of the risks, potential effects, and broader implications that generative AI has on people's cultural practices and the very fabric of people's daily lives. Instead of contributing to debates over the creative or artistic nature of the processes and outputs produced with generative AI, this article aims to delve deeper into the changes that these cultural tools bring to human action. Particularly, this article aims to provide new perspectives on the nature of mediated action by focusing on the intricate relationships and processes between humans and technology in the context of text-to-image generative AI.

This article first presents the theoretical groundwork for mediated action, providing a foundational framework with explanatory potential for conceptualizing how tools influence human thoughts and

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/
Digital Object Identifier 10.1109/MCG.2024.3355808
Date of publication 29 January 2024; date of current version 25 March 2024.

activities. Following this, the article offers a brief overview of the evolution of generative AI, providing a perspective on the cultural and historical contexts that have shaped and facilitated the personal adoption and use of these tools. Building on this theoretical foundation and technological trajectories, the article proceeds to describe and analyze the nature of cocreation between humans and AI, particularly within the context of text-to-image generative AI. To further illustrate the intricate nature of these ongoing transformations, the article discusses potential risks, effects, and ethical concerns related to generative AI. The article concludes with a discussion of some of the changes and implications that generative AI introduces to the realm of mediated action.

THEORETICAL FRAMEWORK

While the emergence of generative AI is frequently discussed in the context of the technological tools and their rapidly improving abilities, it is essential to recognize that the deeper transformations it brings about are inherently socio-technical and socio-cultural, rather than solely technological by nature.[1] Accordingly, the theoretical underpinnings of this article draw from classical sociocultural and cultural-historical theories, rooted in the pioneering work of Lev Vygotsky[2] and his followers.

Sociocultural studies see human action and mind as fundamentally shaped by the "cultural tools," or "mediational means" that individuals and groups employ.[3] Vygotsky proposed that individuals' actions and thinking are profoundly influenced by cultural tools, artifacts, and the interactions they have with other people within specific social activities.[2] Vygotsky viewed "mediated action" as the foundation of all higher psychological processes, describing it as an intricate interplay between the subjects (the actor or actors participating in the activity), the object of their activity, and the tools and artifacts the actor uses to act upon the object.[4] In recent decades, the triangle metaphor for mediated action has also been expanded (Figure 1) to encompass the social world in the form of community, rules, and the division of labor.[5]

Another distinctive characteristic within the sociocultural perspective is the emphasis placed on the relevance of the tools that are available as resources to enhance people's thoughts and actions.[3] Vygotsky emphasized that human interaction with the world occurs indirectly and is mediated by various physical or psychological tools.[6] While he recognized natural language as the primary psychological tool, he also intended his claim to encompass various other means,

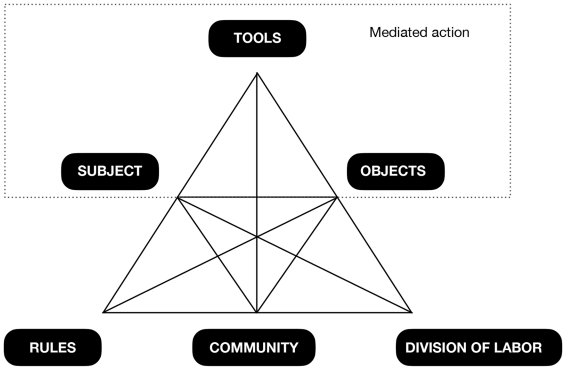

FIGURE 1. Mediated action[4] and expanded activity system.[5]

such as technical tools, works of art, writing, and so forth.[6] The key argument was that cultural tools do not simply mediate or facilitate human actions, but influence and alter the structure of mental functions of the human agent.[6] As Vygotsky was deeply interested in understanding how actions and mental functions are shaped by social interactions and the context in which they occur, this situative nature of social interaction further guides to be attentive to how the entire flow and structure of communicative and collaborative processes, as well as individual mental processes, might undergo transformation in tool-mediated actions.[6]

In general, sociocultural perspectives share some similarities with distributed cognition[7] and its argument that human action and expertise development are material, physically and socially distributed.[8] Human minds, with their limited cognitive characteristics, attain vastly greater power by using various semiotic and physical tools that are major carriers of accumulated knowledge, ideas, and patterns of reasoning of prior generations.[7,8] Furthermore, in many cases cultural tools have been borrowed from distant contexts and shaped mediated action in profound ways in other domains.[9] For example, the advent of computers has fundamentally revolutionized the practices of scientific research by providing computational power, new tools, and methods for knowledge creation and collaboration that stretch beyond the boundaries of local research communities.

Another aspect of distributed cognition is the fusing of tools and minds in social communities and networks into higher level systems.[7,8] Even with constrained individual abilities and cognitive capacities, engaging in collective activities enables shared intellectual efforts, yielding creative resources of a higher quality than what could be achieved by individuals alone.[8] For example, rather than considering creativity as an isolated and solitary endeavor, it has been

argued that radical innovations are more likely to emerge when ordinary minds engage in the creation, development, and use of new and innovative tools and the associated social practices that facilitate extraordinary creative achievements.[8]

From Communities to Collective Actions

While theoretical insights on mediated action and expertise development have been oftentimes associated with professional communities and on-site interactions, the socio-technological developments have also opened the door for various kinds of maker movements and online communities outside workplaces and formal institutions. These communities illustrate how people from various backgrounds have harnessed new tools and computational resources for making, customizing, creating, and sharing ideas and artifacts.[10] It has been argued that rapid socio-technological advancements do not simply involve a shift in the technical infrastructure but also a shift in the social practices.[10] These changes have pointed communities toward a more participatory culture,[10] one in which people are provided with the means to participate in personally meaningful projects and share their ideas, creations, and artworks in communities.[10]

The modern maker movement also embraces sharing of not only the products of design, but the whole process of making with videos, blogs, images, and instructions. This ease of sharing lowers the barrier of entry as a newcomer can build on someone's else's ideas, designs, code, and instructions when making or customizing their own artifacts.[10] Active communities have formed around generative AI, too. When participants mutually share their ideas and intellectual resources for the collective, they also participate in a process that has been called the creation of "collective intelligence"[11] and they follow the "hacker ethic" imperative of sharing information and results with others.

However, individuals and communities can use generative AI to create digital images without a comprehensive or even rudimentary understanding of its mechanics and its impact on mediated action. In this context, people may become unreflective or even unknowing consumers of cultural tools,[3] potentially diminishing their capacity for agentive and informed engagement with technology. As generative AI progressively integrates into numerous facets of daily life in the AI era, thereby influencing people's thinking and conduct, it becomes increasingly important to re-evaluate the role(s) that tools play in cultural practices and mediated action. The following section provides a brief overview of larger trajectories in computing that have paved the way for the inclusion of text-to-image generative models within the toolkit accessible to non-ML experts.

CAPABILITIES OF FOUNDATION MODELS

In the decade spanning from the early 2010s to the present, a sequence of breakthroughs in AI ignited an ongoing renaissance in the field, with implications unfolding into 2023.[12] These breakthroughs signified a paradigm shift from conventional rule-based computing to data-driven computing. The former entails the construction of classical programs to define explicit rules and algorithms for data processing. The latter involves the development of systems that learn optimal behaviors from large datasets. This shift was catalyzed by an unprecedented surge in the availability of diverse high quality training data, coupled with the rapid proliferation of cost-effective and scalable hardware, which proved amenable to the training of AI models.[13] Concurrently, a deeper understanding of the potential use cases of neural networks further facilitated this transition.[13]

A notable milestone in this paradigm shift is deep learning, a technique able to successfully capture and emulate complex cognitive functions.[13] This technique, built upon the framework of a deep stack of neural network layers, has demonstrated remarkable aptitude in domains such as image recognition, speech processing, natural language understanding, and more. It especially excels with media and real-world data that classical programming struggles with, including images, videos, speech, text, and audio. The implications of this paradigm shift in computing are deeply felt in societies across the world, impacting domains such as information and media, pricing of goods, stock trading, policing and sentencing, autonomous traffic, online dating, communication, and smart household devices, to mention a few.[12] In the domain of visual arts, the implications were felt in the early 2020 s with the advent of Dall-E, Midjourney, Stable Diffusion, and many other text-to-image generation systems.

The larger the training datasets and sizes of AI systems trained with those data, the better the systems become at what they do. Notably, as the systems grow, new capabilities emerge that were hitherto absent in smaller scale counterparts. For example, after reaching a certain size, neural network-driven image recognition systems no longer required meticulous manual "feature engineering" for specifying what

to look for in images. Instead, the larger systems would acquire all that from the training data. Another threshold was passed when system sizes reached the level that is called "foundation models"—extremely large neural networks that have been trained with massive amounts of broad, unlabeled data.[1] Foundation models exhibited properties that were unanticipated and unplanned, such as large language models' malleability for a broad range of tasks through prompting, familiar from ChatGPT.[1]

Foundation models challenge the old sociotechnical perspectives of computer systems across several crucial dimensions. First, they distill vast quantities of human-generated outputs into massive neural networks. Second, they exhibit the capacity to generate entirely novel outputs that diverge from their training data, introducing unprecedented machine generativity and innovation—both of course relying on mechanisms that differ greatly from how humans create and innovate. Third, AI systems can enhance their outputs through interactions with other AI systems trained on different datasets, fostering a mutual amplification of capabilities[12] (also users commonly fine-tune their Midjourney AI art prompts using ChatGPT). Fourth, foundation models can be multimodal, trained with any digitizable modes such as text, images, speech, structured data, and raw instrument data. This versatility empowers them to seamlessly integrate and synthesize information in diverse formats.[1] Fifth, post-training, they demonstrate adaptability by fine-tuning to a wide spectrum of tasks and use cases.[1] Sixth, they undertake complex tasks traditionally associated with cognition and long thought to be beyond the reach of computing.[13] Seventh, they generate a loop where AI systems are trained with data from humans and AI, which affects how humans and AI generate data, which are then fed back to the system[12]; in the words of an old adage, "first we build the tools and then those tools build us."

This description of technology—a learning, creating, and communicating, multimodally capable collective of machines that combine the wisdom and creativity of massive crowds of people and can sometimes surpass them—eludes many theoretical descriptions of the role of tools in mediated action. What is more, AI systems are not "just technology" but sociotechnical systems shaped by physical resources, human effort, and societal systems, reliant on substantial human-made inputs.[14] Its development is influenced by prevailing power dynamics, fueled by capital investment, and aligned with established interests. The following section analyzes the impact of generative AI on mediated action.

MEDIATED ACTION AND GENERATIVE AI

Conceptualizing and understanding the complex human–machine systems of the age of AI is profoundly challenging.[12] This section illuminates those complexities by re-examining the fundamental aspects of mediated action within the context of text-to-image generative AI. Consistent with the theoretical approach above, this section proposes that neither the process nor the outcomes of AI-driven image generation are confined within a single entity, but ultimately, both are diffused across extensive networks of human and nonhuman configurations.[15]

Being designed and programmed by humans working in and for corporations, the mechanism and functionalities of generative AI embody human intentions and business models. The architectural design choices made by the programmers, as well as the selection of the massive datasets used for training, are both products of human decision-making, ultimately shaping the behaviors exhibited by the AI algorithm.[12] Thus, these decisions influence the mediated action in which the user engages.

Another important aspect to consider is the training process of the models. These models are trained using vast datasets comprising hundreds of millions of images sourced from the Internet, along with the accompanying texts created by authors, social media users, and others. In addition, many image generation models rely on specific datasets that were manually labeled by humans, and these labeling decisions significantly shape the behavior exhibited by the system.[12] The training data and the manual labeling process profoundly impact the model's output, that is, the images the AI generates in response to different text inputs given by users. Yet, the exact processes of a trained neural network that generate outputs from given inputs are opaque even to those who created these systems in the first place.[12]

The operation and outcomes of generative AI are not predetermined or fixed but emerge through interactions with humans who guide the AI tool through word prompts. That illustrates how AI does not possess a magical ability to create pictures on its own, but it must be used by the human agent. It also demonstrates how writing has evolved to have a powerful mediating impact on image generation, serving as a tool to guide another tool in producing digital artifacts. Such mediated action of producing prompts is also shaped by the tool as the prompt needs to be crafted to communicate effectively with the AI, guiding it to produce images in a certain style, given scene

composition, using certain materials, angle of view, and so forth.

These tools facilitate image creation by externalizing evolving ideas so that users can iterate and refine their ideas and their creations, for example, by modifying their prompts or by making variations. The process of prototyping possible solutions can be iterated many times over until finally reaching a desired outcome. From the sociocultural perspective, mastery of prompting that connects cultural tools in new ways is also shaped by the expertise and the skill of the user in employing these particular means in object-oriented actions.[9] Generative AI can be understood as a tool in mediated actions[9]—but as one that transforms the individual and collective agency of social actors, as well as the structures and social practices that support and constrain agentive actions. These objects can vary from creating illustrations for a PowerPoint presentation to exploring innovative ways to stretch the boundaries of an expert culture, whether that is graphic design, product design or art, to name a few.

Furthermore, the communities around generative AI demonstrate the augmentative effect of collectives suggested by many theorists of participatory culture.[10,11] Text-to-image generative models have become increasingly popular means for creating images and artwork based on text prompts in natural language. The smooth learning curve and low barrier of entry of some of the best-known text-to-image generative AI tools have made content generation readily accessible to newcomers and nontechnical users. Users of text-to-image generative AI have formed communities in various social platforms, in which people share their artworks, prompts, tips, and tricks, as well as engage in various kinds of object-oriented discussions around generative AI.

A case example is the popular image generation tool Midjourney, which can only be used on the social media platform Discord, where each user prompt and corresponding AI-generated image can be seen by everyone on the same channel (as of August 2023, Midjourney's server had 14 million users divided to thousands of channels). In a 2022 interview Midjourney's founder David Holz recollected that the collective design was adopted to facilitate individuals' collective imagination: In early tests, working in separate workspaces failed to expose users to the capacity of text-to-image AI.[16] Holz described how isolated test users would prompt the AI to draw something mundane, like a dog, and if encouraged, perhaps a pink dog, and once they got an AI-generated picture of a dog, "they go 'okay' and then go do something else." But once dozens or hundreds of users were put on the same channel, said Holz, "they'll go 'dog' and someone else will go 'space dog' and someone else will go 'Aztec space dog,' and then all of a sudden, people understand the possibilities, and you're creating this augmented imagination—an environment where people can learn and play with this new capacity."

Accordingly, prompt ideas can be often traced to images and prompts shared by others in communities built around generative AI. Prompts can be borrowed and developed based on the ideas and prompts of others, and then tailored in terms of the specific object at hand. These communities can also be important for expertise development, as they provide informal mentorship, where more knowledgeable others provide resources and support for newcomers in their creative endeavors.[10] As different applications for AI-based content creation continue to emerge, AI communities are also actively exploring new combinations of these tools across modalities, such as text, image, video, and audio. For instance, some are experimenting with creating images based on prompts that were formulated using chatGPT, a language model. Others are delving into the creation of AI-generated videos using images generated through Midjourney, and then adding music to their videos by using Audiocraft, and so on. In all, it illustrates the cross-pollination and emerging convergence of tools, media, and techniques that enables these communities to stretch their creative powers in emergent ways. Yet, the widespread adoption of such tools does not come without complex ethical, legal, societal, and environmental concerns.[1]

UNPACKING THE SOCIETAL AND ETHICAL IMPACTS

As the generative AI era unfolds, so do its complex ethical, legal, and societal implications. Those implications span from copyright infringement and data protection to the perpetuation of biases and existing inequalities, as well as concerns with its impacts on environment, power dynamics, democracy, human agency, and social fabric. This section presents eight key concerns that are becoming increasingly urgent as generative AI becomes integrated in people's daily lives; each concern raising fundamental concerns about what kind of future are those tools helping to shape and for whom.

First, how generative AI is trained challenges copyright, privacy, and ownership. The process of web crawling massive datasets to train large AI models may risk infringing on the privacy and ownership rights of data subjects.[1] Images and associated text, often scraped from online sources, are used without

securing consent from image owners. This process lacks transparency, and fails to provide an option for individuals to exclude their personal pictures from these datasets even when they may contain private and sensitive information.[17] Even intimate photos intended for a limited audience, such as vacation snapshots, wedding portraits, or personal family gathering photos, can become unwitting training material for foundation models without consent and without consideration for their cultural context and significance.

AI's capacity to reproduce the unique styles of individual artists, artistic movements, art history, and pop culture icons has evoked concerns regarding potential copyright infringements. For example, Greg Rutkowski, an artist renowned for his fantasy artwork, found himself in a situation where numerous enthusiastic fans generated tens of thousands of pictures imitating his unique style.[18] This phenomenon also illustrates the risk of genre hijacking, wherein an entire fantasy genre risked becoming monopolized by a single aesthetic. Generative AI excels at emulating existing styles, but it risks content creation stagnating into mere reproduction of existing patterns, impeding diverse, original artistic expressions—perhaps earning it the title "degenerative art" rather than generative art.

The use of generative AI to imitate artistic styles highlights how the rise of new cultural tools also challenges prevailing laws and regulations. At present, existing copyright statutes are unclear on whether the concept of "fair use" can extend to machine learning models, where copyrighted materials are employed without replicating the original images as present in the training dataset.[1] The question of how copyright regulation should adapt raises important legal questions, such as what limits does copyright law set to innovation and progress in AI. It further illustrates how the use of AI-driven content creation blurs the lines between originality, fair use, and creative expression, and thus, forces a reevaluation of existing regulations to adapt to this evolving landscape.

Second, adoption of foundation models risk neutrality, fairness, and equity. The algorithms reflect the values and biases of their creators and users, while also inheriting pre-existing stereotypes and prejudices embedded in their training datasets.[1,14,19] A mounting body of research indicates that machine learning systems may perpetuate or amplify societal biases and negative stereotypes, particularly related to ethnicity, race, gender, and sexual orientation.[1,14] Given that the same training data and model can serve myriad applications, biases may also propagate to many applications and settings.[1]

Third, generative AI opens the door for new unethical uses. User-friendly generative AI tools can be used to create remarkably persuasive fake images, videos, and text, thereby accelerating the spread of fake news, mis- and disinformation, and deep fakes.[1] A prevailing consensus acknowledges the dire threat misinformation and fake news pose to democracy given their potential to influence opinions, voting behaviors, and ultimately, lead to widespread confusion about what is true and real.[1] In addition, the ability to generate high quality synthetic images and videos can be used to embarrass, harass, intimidate, and extort individuals.[1] Even without malicious intent, a deluge of AI-generated content poses new challenges to traditional media and media literacy. While much discourse focuses on AI alignment—ensuring AI adheres to human values and objectives—the risk of human misuse of AI to inflict harm remains a more pressing problem.

Fourth, the emergence of new cultural tools not only reshapes the dynamics of mediated action but also remodels power dynamics.[9] A crucial determinant of power relationships in the AI era is the ownership of data and models.[1] Notably, the development of foundation models has been primarily driven by large corporations, resulting in a highly concentrated market and power. Control over development of AI systems lies in most parts beyond the purview of democratic decision-making, even when those systems carry significant societal risks, spanning job displacement, undermining the social contract, eroded democracy, and societal upheaval. Generative AI gives corporations new tools for behavior engineering and opinion swaying.

Moreover, the question of who possesses the best data and powerful computational resources will likely also dictate who can produce cutting-edge tools and models in the coming years.[1] Anticipating a flood of AI-generated content on the web, the 2020s has witnessed a race to amass the most data not yet tainted by AI outputs. This drive comes from research results showing that training generative models with generated data leads surprisingly quickly the model to collapse into subpar results.

Fifth, much trust is being vested on AI systems even as they are still relatively unreliable. Their widely acknowledged weaknesses—such as opacity, brittleness, and susceptibility to spoofing—are just the tip of the iceberg. Model drift may cause gradual performance deterioration, breaking downstream applications dependent on its functioning. Many foundation models are chaotic systems, characterized by unanticipated emergent behaviors. In some neural network computations graphics processing units work

nondeterministically and cannot guarantee the same result on every run. The more AI systems interact with each other, the more unpredictable those chaotic systems are. Their behavior is not captured by deterministic and reductionist thinking, and explainability is a major challenge to them. When systems based on foundation models fail, accountability and liability for malfunctioning are unclear. For example, does collective intelligence or wisdom of the crowds imply collective responsibility?

Sixth, large AI systems may perpetuate and exacerbate existing global and local inequalities. The way in which AI systems are trained slices the world into classifications that reflect the social and cultural categories of a narrow segment of society, further restricted on a global scale.[14,20] Once trained, the systems amplify the values and beliefs of those whose voices are heard loudest in the training data, leaving the voices of the poorest and the most marginalized unheard. In addition, the ability to automate tasks at mediocre but not exceptional level heightens the significance of domain expertise, yet simultaneously jeopardizes human expertise, skill, and capacity in the big picture. This shift also poses a threat to transforming human agency from active agents to passive users.

Ethical concerns have also been raised in terms of the human labor in machine learning model development. A common harmful practice involves hiring crowdworkers under minimal labor safeguards, exploiting vulnerable individuals such as refugees, incarcerated individuals, or those grappling with significant economic hardship.[14] Moreover, crowdworkers frequently endure working conditions marked by physical and mental strain, stressing their health and well-being.[14]

Seventh, the impact of AI on the human condition is not well understood. AI-human hybrid actions introduce a layer of ambiguity regarding agency. Even as AI systems do not have intentional states or agency, a challenge arises concerning the erosion of ontological boundaries between natural and artifice, not necessarily mirroring reality but in people's minds. This is coupled with forward and reverse anthropomorphization, projecting human attributes to machines and machine attributes to humans.

Eighth, the construction and maintenance of massive AI systems exacts a heavy toll on the environment. As AI applications continue to expand and more powerful models are developed, also their environmental impacts have become under critical scrutiny.[1,14] Training large-scale generative AI models requires substantial computational power, resulting in considerable energy consumption. The energy demands of data centers and high-performance computing systems have raised environmental concerns and sustainability issues.[1,14] In addition, rare earth metals needed for the massive computing systems are typically mined in developing countries, but the countries that benefit the most are not the countries where these metals are extracted.[14] When these devices reach the end of their life cycle, a significant proportion of this waste is also shipped to countries that are far from the recipients of the benefits.

DISCUSSION

This article explored the complex relationship between humans and text-to-image generative models through the lens of mediated action emphasizing the crucial role that tools play in shaping human thought and actions. This article illustrated how generative AI, such as other influential tools, carries a wealth of accumulated knowledge, ideas, and patterns of reasoning from previous generations.[7,8] The foundation models that now underpin tool-mediated activities would never have existed without advances in available training data, computing power, and machine learning techniques,[13] and their utility for a wide range of downstream tasks was largely unanticipated.[1] GenAI's technological trajectory illustrates how cultural tools are oftentimes borrowed or applied from quite distant contexts and this history leaves traces in mediational means and, consequently, in mediated actions.[9] As noted by Wertsch, the introduction of new mediational means, often coming from "outside," has the power to transform existing forms of mediated action, to the extent that one may question whether or to what extent the same form of action is involved at all.[9]

Generative AI is increasingly blurring and redrawing the old boundaries between humans and technology, suggesting that viewing human users as the sole source of creative activities can be problematic, if not misguided. While the objectives, expertise, and creativity of human users are crucial in AI image creation, it is essential to acknowledge that the process involves multiple actors through diverse roles, mechanisms, and objectives. Researchers have spent years developing the AI architectures and training regimes. Programmers working for corporations have customized foundation models for specific purposes. Millions of data subjects have (often unknowingly) contributed data and labels for training the models.[1] Neural networks turn all this human insight into an image-generating machine.

What was once viewed as a one-sided relationship, with the human at the helm, has evolved into a

symbiotic, complex relationship involving not just users and technology but also the creative contributions of millions, the wisdom of the crowds, and collective intelligence. This evolving hybrid complicates traditional perspectives on agency and control. The transformative potential of generative AI tools, as well as the ethics of the emergent social dynamics they generate, remains poorly understood. The deeper one delves into the complexities of generative AI in the context of mediated action, the clearer it becomes that generative AI is reshaping how people interact, think, and work with technology in object-oriented actions. This article proposes four transformations that affect how mediated action is perceived, which are described below.

First, generative AI resurrects an age-old discourse on agency, a debate that spans over six decades, and adds a number of new perspectives. Coined around the same time as the term AI, a parallel but distinct term IA (intelligence amplification or intelligence augmentation) was introduced. One aimed to replace humans in cognitive tasks, while the other sought to enhance human cognitive abilities through human–computer collaboration. Those two approaches inherently generate a tension concerning agency—one approach suggesting some semblance of autonomous agency in the system itself, while the other focusing on human agency with technology in a supportive role.

Over decades, new autonomous systems have continually prompted questions about agency and its locus, from time to time challenging established theories that attribute agency solely to humans within the system. Although AI systems still lack intentionality and agency, in modern systems of individual users, groups of users, and technology, AI systems wield decision-making authority within the loop, users delegate decisions to them, and users communicate and interact with them in new ways. In generative AI, the locus of control is distributed between humans (as initiators of interactions through prompting as well as originators of AI training data) and networks of AI systems (as aggregates of actions of millions of people, based on which AI systems can carry out tasks delegated to them). Those systems are autonomous, capable of decision-making and inference, responsive to multichannel stimuli, and able to communicate with other AI tools. What is more, many foundation models are not restrictively narrow in their applications but can perform a broad range of jobs without a need for retraining, and through in-context learning, few-shot learning, or zero-shot learning can do jobs of which their training data contains no examples.

This intelligence-augmenting dynamic shifts the user's relationship with technology from a passive tool and its user to an active cocreator, leading to decentralized collaboration between many human and AI actors. While users have a lot of freedom in image making, text-to-image AI tools facilitate, guide, externalize, and otherwise mediate the iterative process of prompt engineering and image creation. Generative AI's behaviors, learned from millions of human outputs, blur the distinction of the tool and the human as separate; suggesting a complementary relationship. In terms of distributed cognition,[7,8] cognitive load in those actions is relocated, rebalanced, and redistributed across multiple human/AI actors, which enables one to transcend the human cognitive limits almost imperceptibly.

Yet in this human-AI symbiosis, human-AI cocreation challenges the "I can" mentality. This interaction is not merely instrumental, but it is poised to reshape the way people think and experience the creation process and themselves as creators. As cultural tools go beyond simple mediation or facilitation of human actions—they actually influence and alter the very structure of mental functions—it remains an open question how generative AI tools affect both individual and collective action, as well as perceptions of one's own agency.

Second, generative AI adds another layer of complexity to the already nuanced perception of the "user" as the controller of technology. Within the ceaseless feedback loop of generative AI, the actions of each actor within the interconnected network mutually influence the responses and states—whether cognitive or computational—of other components of the system. In a system where users shape technology that, in turn, shapes users, the relationship between users and technology is nonlinear and emergent, typically involving multiple tools and tool-mediated actions. Historicity, cultural influences, and human sensitivities are ingrained not just in an individual user's interaction, but across the entire user spectrum and in distillations of massive crowds of people. As tool mediated actions become increasingly multiply mediated, involving large-scale networks comprising both human and nonhuman actors, predicting and understanding these emerging interactions and their consequences becomes ever more challenging. Moreover, the opaque nature of foundation models, along with industrial secrecy and legal protections of intellectual property, add to the difficulty of grasping these emerging interactions and their outcomes.[1,12]

Prompting has become a favored mode of human–computer interaction in the realm of generative AI. However, conversational interfaces, especially in the context of multimodal AI models, stretch the

conventional conceptions of human–technology interaction, especially as they can respond to a broad range of media forms, including music, speech, pictures, video, and more, not just text. To use these new tools effectively and efficiently, individuals must adapt and learn new practices, skills, and ways of reasoning.[9] The widespread adoption of generative AI is accelerated by its ability to provide easy access to supercomputer-level performance, enabling users to effortlessly achieve unprecedented results. This allows users the ability to externalize their ideas in natural language, which Vygotsky considered the primary psychological tool. While natural language prompts remain the most common method for controlling generative models, this approach has serious caveats due to the way conversational systems are trained with data.

One of the most serious data-related caveats is bias. All kinds of data that web crawlers harvest from the Internet—text-image pairs, for example—are used to train generative models.[1] As a consequence, achieving optimal prompts for generative AI systems requires mastery in how to articulate one's desires—whether for music, sound, images, 3-D objects, and so forth—in a way consistent with the text descriptions in the training data. But these training datasets are tainted with various biases, and how the training data slice the world into categories reflects the values and perspectives of those who contributed the data: Typically a narrow, affluent segment of population in industrialized countries.[14,19] The voices of the poorest and the most marginalized often go unheard.[21] As users need to learn to describe the world and verbalize their needs in the same way the contributors do, it raises complex questions of how human actions and mental functions are shaped by values and cultural practices of the few.

What is more, the relationship between the user and technology is not just interactional, but can be invisible or evolve into an intangible connection involving sensing, predicting, emotional resonance, and even a sense of partnership. AI systems that, based on billions of earlier interactions, can predict what users need before the user even acts require revisiting how the action/reaction relationship is seen. AI systems may communicate with each other more effectively and efficiently in languages or protocols that elude human understanding, thereby operating in a realm beyond human understanding.

Third, the generative AI tools that are used for creative purposes stand apart from most classical tools in a profound way: their outputs defy predictability. Yet, that is not a bug but rather a desired feature for fostering creativity. Whereas many tools of the 1900s were deterministic, predictable, and reducible, many AI-based systems—especially generative models—are not. Their outputs eschew determinism, producing slightly different results with each iteration. Their behavior defies reductionism, exhibiting emergent patterns that cannot be reduced to single parameters of the model. Consequently, tool-mediated actions do not ensure specific outcomes. Moreover, tools that can predict user actions based on historical behavior, biometric data, or actions challenge the conventional action–reaction relationship between user and the tool.

Fourthly, AI systems are progressively re-engineering, often with mixed outcomes, the social dynamics among humans in the system. Even if the generative AI "wisdom of the crowds" can suffer from pronounced, undesirable biases due to the lack of the voices of the poorest and the most marginalized, these systems still exhibit collective intelligence at many levels. They distill the creative contributions of countless individuals into very large models. In many instances, their users form open-source communities that share prompts, outputs, techniques, and tips and tricks.

In summary, the concept of mediated action provides valuable insights into the intricate interplay between humans and technology. However, the advent of generative AI compels researchers to reevaluate the role(s) that tools play in cultural practices (see Figure 2). As AI tools, material artifacts, and the social fabric become increasingly interwoven, AI systems also transform and mediate human actions in countless ways. Nonetheless, the opaqueness of foundation models, along with industrial secrecy and protected intellectual property, creates substantial challenges in predicting and understanding these emerging interactions and their implications.[1,12] Moreover, these activity systems of humans and AIs are in constant flux, which makes the task of anticipating and analyzing their emergence significantly more challenging in the age of AI.[12] Even in the absence of comprehensive theories of how social and cultural practices are affected by pervasive AI,[12] there is a pressing need to understand how these emergent relationships form and transform, how human agency is constituted in networks of human and nonhuman actors, how both humans and machines adapt their individual and collective behavior in these interactions, how new forms of power relationships and social structures shape and constrain agentive actions, and what the short- and long-term consequences of such transforming relationships are. This makes understanding the myriad, overlapping forces that shape human agency in the AI era an urgent priority for future research.

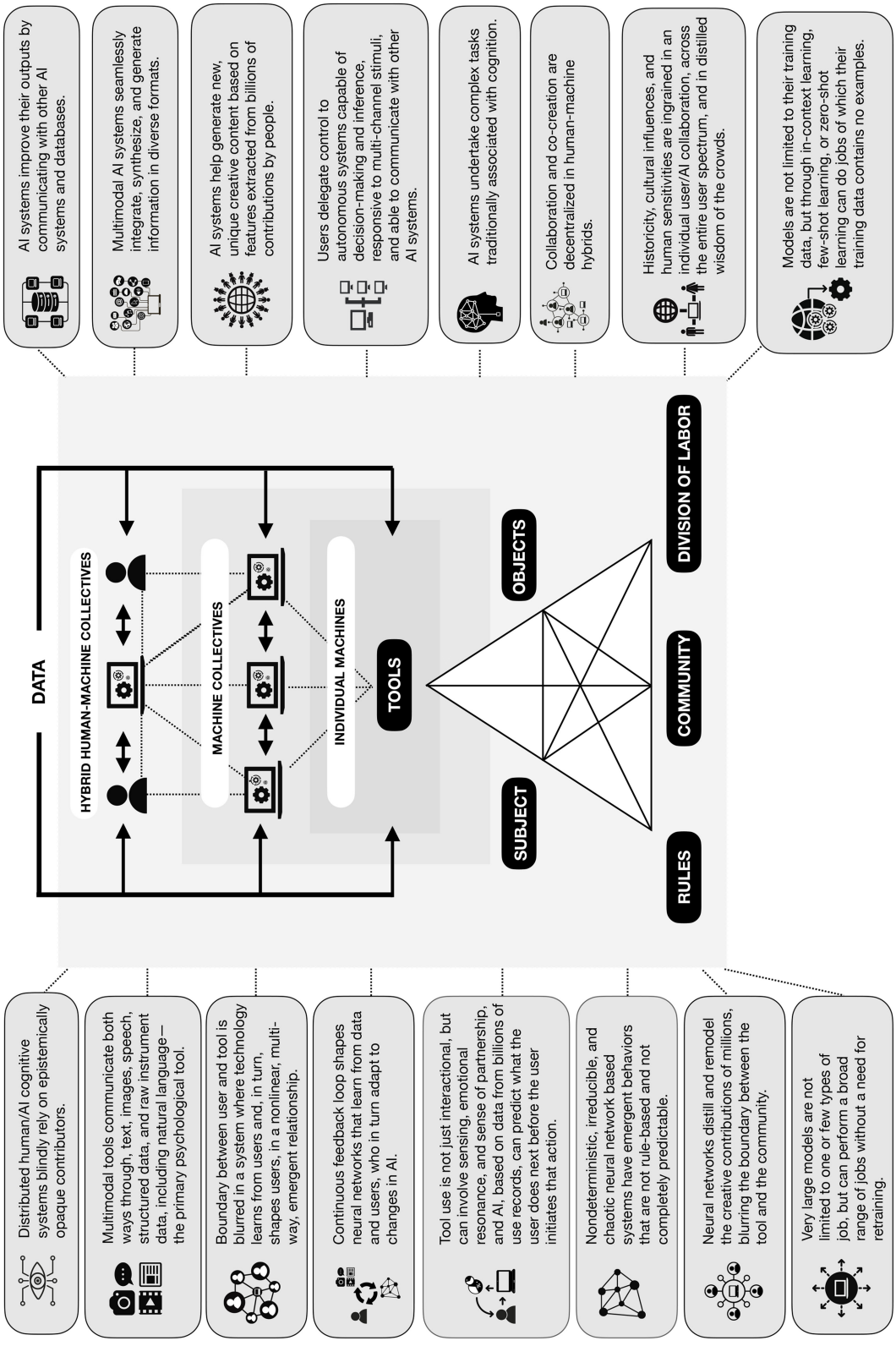

FIGURE 2. Characteristics of foundation models in the context of generative AI and in relation to the theory of mediated action.

ACKNOWLEDGMENTS

This work was supported by the Strategic Research Council (SRC) established within the Research Council of Finland under Grant #352859 and Grant #352876. The authors would like to thank January Collective for core support.

REFERENCES

1. R. Bommasani et al., "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.
2. L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA, USA: Harvard Univ. Press, 1978.
3. J. V. Wertsch, *Voices of the Mind: Sociocultural Approach to Mediated Action*. Cambridge, MA, USA: Harvard Univ. Press, 1991.
4. M. Cole and Y. Engeström, "A cultural-historical approach to distributed cognition," in *Distributed Cognitions: Psychological and Educational Considerations*, G. Salomon, Ed., Cambridge, U. K.: Cambridge Univ. Press, 1993, pp. 1–46.
5. Y. Engeström, "Innovative learning in work teams: Analyzing cycles of knowledge creation in practice," in *Perspectives on Activity Theory*, Y. Engeström, R. Miettinen, and R.-L. Punamäki, Eds. Cambridge, MA, USA: Cambridge Univ. Press, 1999, pp. 377–404.
6. J. V. Wertsch, "Mediation," in *The Cambridge Companion to Vygotsky*, H. Daniels, M. Cole, and J. V. Wertsch, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2007, pp. 178–192.
7. R. D. Pea, "Practices of distributed intelligence and designs for education," in *Distributed Cognitions*, G. Salmon, Ed. New York, NY, USA: Cambridge Univ. Press, 1993, pp. 47–87.
8. K. Hakkarainen, "Mapping the research ground: Expertise, collective creativity and shared knowledge practices," in *Collaborative Learning in Higher Music Education*, H. W. Helen Gaunt, Ed. London, U.K.: Routledge, 2013, pp. 13–26.
9. J. V. Wertsch, *Mind as Action*. Oxford, U.K.: Oxford Univ. Press, 1997.
10. H. Jenkins, R. Purushotma, M. Weigel, K. Clinton, and A. J. Robison, *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century* (The John D. and Catherine T. Mac Arthur Foundation Reports on Digital Media and Learning Series). Cambridge, MA, USA: The MIT Press, 2009.
11. P. Lévy, *Collective Intelligence: Mankind's Emerging World in Cyberspace*. New York, NY, USA: Plenum Press, 1997.
12. I. Rahwan et al., "Machine behaviour," *Nature*, vol. 568, no. 7753, pp. 477–486, 2019.
13. A. Darwiche, "Human-level intelligence or animal-like abilities?," *Commun. ACM*, vol. 61, no. 10, pp. 56–67, 2018.
14. K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT, USA: Yale Univ. Press, 2021.
15. H. Vartiainen, P. Liukkonen, and M. Tedre, "The mosaic of human-AI co-creation: Emerging human-technology relationships in a co-design process with generative AI," no. G5FB8, 2023. [Online]. Available: https://doi.org/10.35542/osf.io/g5fb8
16. J. Vincent, "'An engine for the imagination': The rise of AI image generators," *The Verge*, Aug. 2022.
17. E. S. Jo and T. Gebru, "Lessons from archives: Strategies for collecting sociocultural data in machine learning," in *Proc. Conf. Fairness, Accountability, Transparency*, 2020, pp. 306–316.
18. M. Heikkilä, "This artist is dominating AI-generated art and he's not happy about it," *MIT Tech. Rev.*, vol. 125, no. 6, pp. 9–10, Sep. 2022.
19. R. Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, MA, USA: Polity, 2019.
20. G. C. Bowker and S. L. Star, *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA, USA: The MIT Press, 2000.
21. V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY, USA: St Martin's Press, 2018.

HENRIIKKA VARTIAINEN is a university lecturer and senior researcher with the University of Eastern Finland, School of Applied Educational Science and Teacher Education, FI-80101, Joensuu, Finland. Her research interests include data agency, co-design, design-oriented pedagogy, and the role of AI in art, craft, and design. Vartiainen received her Ph.D. degree in education from University of Eastern Finland and the title of Docent in Education from University of Jyväskylä, Finland. Contact her at henriikka.vartiainen@uef.fi.

MATTI TEDRE is a professor of computer science with the School of Computing, University of Eastern Finland, FI-80101, Joensuu, Finland. He is the Principal Investigator of a large, six-year national project that studies how to empower learners with insight into the mechanisms, opportunities, and dynamics of data-driven (AI) systems, but also their weaknesses, biases, and risks, and how they can be misused to discriminate, polarize, create insecurity, and break trust. Tedre received his Ph.D. degree in computer science from the University of Joensuu. He is a member of IEEE and ACM and the corresponding author of this article. Contact him at matti.tedre@uef.fi.