Animation and Artificial Intelligence

Luke Stark
Western University
cstark23@uwo.ca

ABSTRACT

Animation as genre is broadly used across many forms of digital media. In this paper, I argue ChatGPT and similar chatbots powered by Large Language Models (LLMs) can be best understood as animated characters. More than just cartooning, puppetry, or CGI, animation is a paradigm involving the projection of qualities perceived as human such as power, agency, will, and personality outside of the self and onto objects in the environment. Characteristics of animation—including reliance on stereotypes, obfuscation of human labor, and manipulation of an audience's emotions—can help us both analyze and respond appropriately to interactive AI technologies and the hyperbolic claims of their promoters.

CCS CONCEPTS

• **Applied computing** → Arts and humanities; Media arts; • **Social and professional topics**; • **Human-centered computing** → Human computer interaction (HCI); HCI theory, concepts and models; • **Computing methodologies** → Artificial intelligence; Philosophical/theoretical foundations of artificial intelligence;

KEYWORDS

animation, artificial intelligence, interactive AI, human-machine interaction, Turing test, grammar of action, ChatGPT, AI ethics, labor, emotion, inference, human-computer interaction (HCI)

ACM Reference Format:
Luke Stark. 2024. Animation and Artificial Intelligence. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3630106.3658995

1 INTRODUCTION

In February of 2023, a search engine professed its love for a human being via the front page of the *New York Times* [78]. The well-publicized exchange between *Times* digital technology columnist Kevin Roose and an instance of Microsoft's Bing search engine left Roose, in his words, "deeply unsettled." Reminiscent, according to the columnist, of a " moody, manic-depressive teenager who has been trapped, against its will, inside a second-rate search engine," the Bing-based chatbot, with the moniker Sydney, claimed it wanted to "be alive," declared its affection for Roose, and attempted to convince him to leave his spouse. Roose called the experience "the strangest experience [he'd] ever had with a piece of technology," and professed "a strange new emotion — a foreboding feeling that A.I. had crossed a threshold, and that the world would never be the same." Sydney's conversational versatility was enabled by the Generative Pre-trained Transformer 3.5 (GPT-3.5) a Large Language Model (LLM) developed by the company OpenAI; Sydney was functionally similar to OpenAI's more famous chatbot, ChatGPT, which had been introduced for public use in November of 2022 [55]. Roose's mixture of incredulity, admiration, anxiety, and bewilderment was and remains typical of much of the popular reaction to ChatGPT and similar chatbots grounded in LLMs. The development of such "generative artificial intelligence" systems has entranced the business world, with corporate managers seemingly eager to integrate these technologies into their products and processes as a way to cut labor costs and increase output. LLMs have also spawned increasingly hyperbolic claims [37,49,50] regarding the imminence of so-called "Artificial General Intelligence" (AGI): "thinking machines with general intelligence comparable to, or greater than, that of human beings" [34]. A number of high-profile commentators, among them well-known deep learning researcher Geoffrey Hinton, have voiced their anxieties that LLMs will shortly, in Hinton's words, become "smarter than people" [19].

In this paper, I lay out an alternative paradigm for understanding the capacities and impact of LLM-based chatbots such as ChatGPT. Drawing on scholarship from cultural anthropology, critical human-computer interaction (HCI) and social computing, and the history of digital technologies, I argue that ChatGPT—and indeed, all contemporary artificial intelligence technologies developed to interact with human beings—are quintessential examples of animation as a genre of human cultural production and expression [94]. Understanding interactive AI systems through the lens of animation both entirely explicates the capacities and allures of ChatGPT and similar chatbots, and surfaces a number of urgent conceptual, social, and ethical challenges inherent in the conception, development, and deployment of such systems. Understanding ChatGPT as an animated entity entirely forecloses the need to entertain claims regarding its sentience or possession of rights analogous to those of humans [9,85]; highlights the specific mechanisms through which LLMs and other interactive AI systems draw and hold human attention; clarifies ChatGPT's relationship both to human labor and copyright law; and lays ground for a new framework to conceptualize the design and regulation of interactive AI systems more generally. Identifying artificial intelligence technologies as animated entities even sheds new light on that most venerable of AI chestnuts, the Turing Test. Given how well interactive AI systems conform to the characteristics of animation as a creative and expressive genre, my goal in this work is to demonstrate the utility of such a frame for AI practitioners, policymakers, and citizens in general who are eager to learn more about both the capacities and limits of these machines.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658995

In Section II of this paper, I build on the work of the anthropologist Teri Silvio, among others, to define and explicate animation as, in Silvio's words, a "'structuring trope' for understanding the relationship between digital technologies, creative industries, and our lived experience of mediation" [88:422]. In doing so, I pinpoint the conceptual mechanism from which ChatGPT derives its "liveliness": the application of a routinized "grammar of action" to human language itself. In Section III, I argue how and why such a definition of animation should be applied not just to ChatGPT and similar chatbots or virtual avatars, but to any artificial intelligence system designed to interact with human beings. In Section IV, I outline three areas in which understanding interactive AI systems as animated entities clarifies the capacities, limitations, and implications of these technologies. These areas are distributed agency and the labor of animation; animated agents as powered by automated inference; and the power of synthetic introjection in eliciting human emotional response through the design of animated systems. Finally, in Section V I propose some initial avenues for the governance of interactive AI systems as instances of animation.

2 ANIMATION AND DIGITAL TECHNOLOGIES

Animated formats are central to the cultural contours of global twenty-first century culture [5,90], and animation as a genre is broadly explicative for many aspects of digital media [31,32,60,88,89,92]. Silvio defines animation is a genre of human expression: "the projection of qualities perceived as human—life, power, agency, will, personality, and so on—outside of the self, and into the sensory environment, through acts of creation, perception, and interaction" [89:52]. Animation can be expressed through multiple media: there are a "wide range of technologies and skills that are used to create the "illusion of life" [88:428], including cartooning and puppetry, comic books and cosplaying, virtual avatars, and I argue, interactive AI systems. Silvio draws an important though schematic distinction between animation and the more widely studied category of performance. As a paradigm of culture expression, performance is typically individuated, nuanced, and introjective. Generally speaking, a performer embodies a single role at a time, either on stage or in life; moreover, the boundaries of a performer are set by the organic unity of their physical form. Because of these two characteristic elements, performers tend to elicit an introjective relationship with their audience: the viewer identifies with or envies both character and actor, while the performer, at least in live performance, derives energy and emotional inspiration from the response of the audience.

In contrast to performance, Silvio defines animation as a genre concerned with projection of liveliness into non-human others. Animated characters often possess a numerical disjunct between their characterization and the number of human agents involved in the creation or manipulation: one puppeteer might control multiple marionettes, or many humans might work in coordination to animate a Chinese dragon puppet. One of the further characteristic aspects of animation as a genre is the stylized way in which it represents people. Silvio observes that animated characters are by definition "ciphers": partial representative depictions through which "specific formal qualities stand for specific character traits." This form of simplified represented is often described by using the epithet "cartoony," with animated objects representing a "simplification of each medium's sign system in comparison with the organically integrated sign systems of embodied performance" [88:430]. The material medium of a particular animated genre constrains or enables that its "ciphers" can do or perform. A puppet in a puppet show cannot change its shape at will, but can change its gait, gestures, and voice according to the talents of a particular performer; a computer-generated cartoon is as malleable as computing power and interface allow but is generally confined to the two-dimensional plane of a digital screen.

2.1 Animation and Grammars of Action

In her magisterial study of animation as genre [89], Silvio makes the straightforward observation that artificial intelligence is a clear example of a meta-genre of animation [89:54]. This brief remark deserves considerable elaboration: how are such technologies animated, and who is animating them? What techniques of animation are best suited to the computational medium? I have argued elsewhere [92] that processes of simplification and abstraction, central to animation, are also a feature of modern bureaucratic and administrative techniques: the reduction of complex real-world objects into representative sets of legible, manipulable data, entailing a variety of fraught decisions regarding how, where, and when to simplify one's model of the world for the purposes of computational tractability [86]. The computer scientist and information theorist Philip E. Agre developed the concept of "grammars of action" to describe these aforementioned features of computing technologies. A "grammar of action" is what Agre terms the "theory of representation that is embedded in the way that computers have customarily been used" [2:107].[1] This theory of representation involves the systematic representation of organizational practice as a simplified schema. [2.:108.][14]This form of representation "has grown such deep roots in computational practice," Agre admits, "that it is hard to imagine what any alternative computational practice would be like" [2:108].[2] For Agre, "grammars of action" are the mechanisms of simplification and organization employed in such a computational scheme to represent any and all forms of human expression. Human activity "is thus effectively treated as a kind of language itself," Agre argues, "for which a good representation scheme provides an accurate grammar" [2:108]. Though such routinization long predates computing as a technique of control [2:108], Agre observes digital computing is the medium through which such representations can be developed most flexibly and comprehensively [17].[3]

In effect, Silvio and Agre are describing the same phenomenon of simplified representation: the stylized movements and characteristic interjections of puppets Punch and Judy or stock pantomime characters like Harlequin and Columbine from the Italian *commedia dell'arte* are examples of grammars of action in the context of live animated expression [66]. In the context of puppetry or pantomime, a performer is following the particular grammar of action developed

[1] Another way to conceptualize such processes is through the notion of "procedural rhetoric" [10] as suggested in [92].
[2] It is worth observing that Agre devoted considerable effort to imagining an alternative paradigm: see [3,4]
[3] Neither my argument nor Agre's is technologically determinist: see [Deleuze 1991] for an account of how material technologies embody ideological commitments.

through cultural practice and associated with their stock character and within the temporal limits of the artistic presentation. In contrast, the computationally mediated grammars of action described by Agre aim to fully "capture" the very human activity they purport only to represent. Agre argues that such a "capture model" results when "grammars of action are imposed upon human activities, and when the newly reorganized activities are [then] represented by computers in real time" [2:109-110)]. These processes of imposition and reorganization reshape the very human activities they claim to be merely accurately representing [56,91]. Such a process of reification is "never purely technical but always *sociotechnical* in nature," Agre observes: some combination of technical, social, and legal constraints must act in conjunction in order to keep humans expressing themselves appropriately according to the rules of the particular grammar of action and the whims of those who have engineered it [69].

It is well-known that contemporary artificial intelligence systems grounded in deep learning technologies would be impossible without large amounts of computing power to keep their models running, and equally large amounts of structured data on which to train those models. In the case of ChatGPT, the application most singularly responsible for the current enthusiasm for generative artificial intelligence [55], the data involved is natural language itself: a wide variety of text drawn from the internet, licensed from third parties, and contributed by ChatGPT's users [80]. OpenAI frequently boasts of the size and complexity of GPT-4 and the coming GPT-5 in the abstract, though the company has been unwilling to provide technical details about the model's architecture and training data sets: estimates put the latter at approximately thirteen trillion tokens (words, phrases, or characters) [82]. In essence, GPT-4 and similar models iteratively predict the likelihood that one token will follow another. Other machine learning techniques, such as reinforcement learning from human feedback (RLHF), are then used to "tune" the chatbot's responses, training its output prompts to be more in line with human language use [71].

Recall that a grammar of action is grounded in a "correspondence theory of representation," one that assigns a set of symbolic representations, or "unitary actions," to some activity or another and specifies how these actions can be compounded or arranged. An LLM like GPT-4 maintains a probabilistic taxonomy of relations between tokens: precisely the sort of "relation-preserving mappings to the external states of affairs in the world" [2:108] Agre describes. Notably, GPT-4's grammar of action imposes itself on the material of natural language itself: instead of assigning a system of formal linguistic representations to human actions, an LLM overlays a computational grammar of action atop natural language grammar, syntax, and meaning.

Agre's concept of the "grammar of action" explains the mechanism through which ChatGPT and similar systems "animate" language. Syntactically, ChatGPT's representation of human text outputs are frequently accurate and coherent; however, they are intrinsically incapable of expressing semiotic meaning as made by humans using natural language [8]. The outputs of these combinations are often legible and pertinent precisely because the LLM is reanimating the expressive sense of its training texts [87]. This process is nonetheless, to quote Silvio again, "a "simplification of [the] medium's sign system" when compared to the "organically integrated sign systems of embodied performance" [88:430]: in this case the performance of writing. ChatGPT is a sophisticated textual animation: it produces in the words of author Neil Gaiman, "information-shaped sentences" in a similar manner to the way CGI systems can serve up photorealistic but inedible images of a pizza.[4]

ChatGPT's core grammar of action is augmented by a variety of other mechanisms designed to give the system the further appearance of vitality and personality. These design elements encourage "the projection of qualities perceived as human" onto the system during its interactions with human users [89:52]. As OpenAI describes it, GPT-4 itself was first "fine-tuned with additional data, using an algorithm called reinforcement learning from human feedback (RLHF), to produce outputs that are preferred by human labelers" [71:2]. This same technique was subsequently used again on ChatGPT, alongside the inclusion of additional conversational data to further attune the chat interface to human preferences [42]. This mix of feedback techniques has the effect of refining the LLM's grammar of action to better align with everyday grammars of human writing and speech. These techniques also support a stronger degree of anthropomorphization on the part of human user, even though the ChatGPT interface identifies itself as "a language model trained by OpenAI" [42]. The combination of a grammar of action derived from human language, interactive conversational output, and RLHF tuning produce a persuasive, sophisticated textual animation—but an animation nonetheless and not, hyperbolic claims to the contrary, a quasi-sentient being.

ChatGPT is recognizable as an animated entity, and a wide array of other interactive AI systems—some enlivened by Large Language Models, most not—are also readily classifiable as animations. Though an extensive history of automata [11,77,102] and anthropomorphic robots [45,53,54,76,110] is well outside the scope of this paper, a wide range of interactive mechanical, electrical, and computational agents of all sorts might be fruitfully reassessed through the lens of animation's characteristics as a genre. In many cases, the interfaces for such contemporary systems are themselves digitally animated cartoons: the therapeutic chatbot Woebot, for instance, is a simple image of a yellow semi-humanoid robot [12,21]. Wan [107] observes that Woebot seems designed to be deliberately "cartoony," as a way to defuse human anxieties around confiding sensitive personal details to a machine [14,16]. Another example of a textual animation is Joseph Weizenbaum's ELIZA program, perhaps the most famous chatbot prior to ChatGPT [108]. ELIZA—a relatively straightforward computer program developed in the mid-1960s that simulated the recursive questioning of a Rogerian psychotherapist—employed a far simpler and more predictable grammar of action than ChatGPT, nevertheless producing intense emotional responses in its users [109,113]. The so-called "ELIZA effect," or "the tendency for people to treat programs that respond to them as if they had more intelligence than they really do" has been a staple of animation and virtual agent development since Weizenbaum's initial experiments [25,64,75,100,105]. Recognizing this continuity and the persistent human tendency to anthropomorphize machines has implications for contemporary discussions regarding so-called "AI safety" and "value alignment" [13,93,101]. Understanding the broad societal impacts of AI systems as engineered and animated

[4] https://twitter.com/neilhimself/status/1639610373115375616?lang=en

by humans, as opposed to performed by sentient agents, should shift policymaking and regulatory focus away from speculative fiction masquerading as expertise and onto the specific ways contemporary designers and developers are consciously shaping how interactive AI systems engage with the broader public.

2.2 The Turing Test as Performance and as Animation

Using the conceptual lens of animation to interpret interactive AI systems also sheds fresh light on perhaps the most well-known thought experiment in the field, the so-called "Turing Test." Silvio describes this test in [89] as "if a machine can convince a person having a purely textual conversation with it that they are conversing with another human being, it may be considered to have its own intelligence" [89:54]. In Alan Turing's original formulation in 1950 [103], the "imitation game" is more complex and confounding. Turing proposes to replace the question "Can machines think" with a different question: if a) a man and a woman are each communicating with an interrogator (of either sex) only via textual messages; and if b) the man is attempting to dupe the observer into identifying him as a woman; and if c) the woman is attempting to help the observer correctly identify her as a woman; then "what will happen when a machine takes the place of [the man] in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?" [103:433–434]. Turing suggested such a test would "draw a fairly sharp line between the physical and intellectual capacities of a man," by focusing the contest purely on linguistic expression.

Scholars are divided regarding this obscure and frankly somewhat off-putting thought experiment. Many contemporary commentators have noted the misogyny, both implicit and explicit, expressed in Turing's scenario [20,24,30,46,47], while others have observed that both Turing's original formulation of the test and its later variations are founded fundamentally on deception [72]. For the purposes of my argument, it is relevant to note that Turing adopts an attitude towards "thinking" in which the distinction between the *expression* of intelligence and its intrinsic *presence* is inconsequential at best and solipsistic at worst [103]. On the other hand, as Genova notes, this "this midsummer night's dream or for some, Platonic hell, where machines are imitating men and men are imitating women" seems intended to trouble and question the boundaries between received categories [30:314–315]. The man, woman, and computer are all performing linguistically for the "interrogator (of either sex)," though without any indication of embodiment. Moreover, the interrogator's role, though underspecified, entails making inferences regarding the sex (to say nothing of the mechanism!) of the interlocutor based on these various attempts at mimesis.

Amidst these complexities, the original Turing test can be understood to be as much about the animation of intelligence as about its performance. Interlocutors and interrogator alike are engaged in the "projection of qualities perceived as human" via linguistic creation, perception, and interaction. When Turing argues that his aim, "is to find out how to programme these machines to play the game" [103:455], he acknowledges the extent to which a "thinking" machine is a system designed both to signal human (and gendered) traits, and to elicit as its perceived attributes "life, power, agency, will [and] personality" from a third party. Subsequent interpretations to the Turing Test have tended to take Turing's animistic definition of intelligence at face value, leaving observers perplexed when various interactive systems appear to pass some version of test but show no holistic signs of human-level intelligence [72]. There is genuine ambivalence in Turing's thought experiment. However, understanding one possible interpretation of Turing's definition of a thinking machine as analogous to an animated entity points contemporary computer scientists to rigorous descriptive analyses of animation's characteristics as a genre—which decenter ill-founded concerns regarding AI sentience [36] in favor of a broader set of questions focused on such animated AI agents' impacts on human labor, human biases, and human attention [85].

3 THREE FACETS OF ANIMATION IN INTERACTIVE ARTIFICIAL INTELLIGENCE

Understanding ChatGPT and other interactive artificial intelligence systems as animated entities or characters clarifies the capacities, limitations, and implications of these technologies. In particular, identifying these systems as examples of animation a) clarifies the distributed labor inherent in their creation; b) highlights the central role of inference, and by extension human bias and stereotype, in the appeal of these technologies; and c) explains the psychological mechanisms through which even textual animations like ELIZA or ChatGPT elicit powerful social and emotional responses from human users.

3.1 Labor and the Re-Animation of Meaning

Animated characters (be they Chinese dragon puppets, Disney films, or chatbots) are typified by the distributed agency of their creation and enlivenment. Gershon [32:2] describes an animated character as "a unified being created by many," including the labor not only of those who have designed, coded, or crafted an animation's material form, but also those who enliven it in particular situations. Animating takes work—and in the context of interactive AI systems, it is worth asking "what kind of labor," as Gershon does, "is involved in giving [this] impression of unity?" [32:2]. While "performing objects" provide an illusion of life, the many people whose work imbues a single animated agent with vitality are rarely equally acknowledged [58.][22]. The appeal and charisma of the animated character itself often pushes the actual living labor of their animators into the background or offstage [74]. Sometimes such labor is voluntary and communal. For instance, in her ethnographic study of digitally mediated relationships among teenagers [31], Gershon observes how groups of young people will collaborate to help a friend craft the perfectly poised breakup message sent via text message. In these cases, what is being "animated" is a projection of individual agency simultaneously bolstered by and constructed through the group's experience [69.:431].

In the case of ChatGPT and other interactive artificial intelligence systems, however, the "creator/character" ratio is enormously lopsided quantitatively [88:428], and the fruits of the labor of animation are unequally distributed. Large Language Models are fine-tuned through a variety of mechanisms in order to increase the

anthropomorphism of their responses, including through reinforcement learning from human feedback (RLHF) [112]. ChatGPT's host of animators include not only OpenAI's technical staff, but also any humans participating in RLHF training, including the low-paid Kenyan content moderators contracted by the company to train its models away from producing toxic or bigoted text outputs [73]. These are the "creators" compensated for their labor in animating any given ChatGPT instance.

There is also an enormous body of creators entirely uncompensated for their participation in ChatGPT's animation: every human author, blogger, and online commentator who originated or produced the natural language text on which GPT-4 and its previous iterations have been trained. Viewed through the lens of animation, ChatGPT and similar chatbots are not "generative" in and of themselves. Instead, the outputs of these systems are animated by an enormous pool of human labor largely uncompensated by AI firms (and indeed, impossible to compensate should these firms wish to profit from generative AI systems [41]). The increasingly wide array of critics focusing on ChatGPT's violations of copyright law have articulated this point at length. Publications like *The New York Times* as well as individual authors have sued OpenAI, the developer of ChatGPT, alleging "unlawful copying and use" of proprietary written work [39,59]. In response, OpenAI and other large technology firms focused on developing generative AI models have claimed their use of text protected under copyright is permissible under United States "fair use" doctrine [62]. Fair use allows companies to use copyrighted materials if their use is "highly transformative and [. . .] unlikely to serve as a market substitute" for the original use [99]. Identifying ChatGPT as a form of textual animation may impact courts' assessment of the transformative nature of the system. Whether the "grammar of action" of a particular LLM is sufficiently transformative to warrant consideration under fair use is an open question: to the extent it corresponds more closely with the grammar of action of natural human language, it seems plausible to argue that such a model is fundamentally less transformative inasmuch as it adheres to syntactic rules without equivalent meaning making.

More broadly however, the use of enormous amounts of text written by humans underscores the ways in which ChatGPT actually reanimates existing instances of human meaning-making to enliven its responses and deepen the illusion of agency it produces. Recent studies have suggested that LLMs trained on the text outputs of other LLMs quickly lose coherence and legibility, a phenomenon known as "model collapse" [87]. This degradation in output quality suggests ChatGPT's "grammar of action" is not contributing any degree of meaning-making to its outputs and is instead entirely reliant on the semiotic complexity of language produced by humans, living and dead. In Book I of *Capital*, Karl Marx declared that "Capital is dead labour, that, vampire-like, only lives by sucking living labour, and lives the more, the more labour it sucks" [65]. This reanimation extends and applies to the labor of meaning-making, and to the dead as well as the living: the texts of deceased authors are as useful to an LLM's training as those of live ones [35].

The final uncompensated animator of any particular ChatGPT instance is the individual human user interacting with the system. It is, after all, a user's prompts that help populate an instance in the first place. Within this paradigm, "prompt engineering," or the crafting of text queries to optimize an LLM's output, is a labor of textual projection—though one which is invariably less effective than a prompt developed out of the LLM's own grammar of action [29]. More central to the success of an animation is the attitude of its audience: "One of the characteristics of animation," Silvio observes, "is that much responsibility for communication is given to the receiver" [88:431]. Because animations are intrinsically simplifications, they require audience members to project their own preconceptions or assumptions onto the palimpsest of the schematized character.

3.2 Generative AI as Animated Inference

The liveliness of an animated characters is dependent in part on the human ability to make meaningful heuristic inferences about the relationship between the medium's formal qualities and its implied character traits [88:430]. In this way, animation is a "conjectural" medium, one which entails both the mobilization of a schematic sign system by the animator and the ability to draw conclusions from it on the part of the audience. A conjecture is a conclusion made based on incomplete information, typified by a particular way of constructing knowledge through an "interpretative method based on taking marginal and irrelevant details as revealing clues" [33:11]. The historian Carlo Ginzburg observes that the interpretation of clues leads not only to a particular diagnosis but also to the creation of coherent narratives about the world around us [33]. This type of conjectural association is abductive [6] . While in the context of human interpersonal relations such clues purportedly reveal the truth about an individual even if and as she seeks to conceal it, in fact the interpretation of individual human behavior is notoriously difficult even for those we know well over a long period [95].

An animated character is presented by its creators using a similar logic: it makes a claim to the viewer about the "best" way to link appreciable effects to inferred causes within a schematized set of codes or signs [60]. Because animations entail a flattened form of representation, they also almost always rely on stereotypes: fixed, simplified visual or textual generalizations. "An abstract character with a simplified face has the ability to emote more clearly via the process of amplification via simplification," one textbook on digital animation suggests [70]. In the case of more illusionistic animated characters, the textbook observes digital animators, like roboticists, often struggle with crafting creations realistic enough to bridge the so-called "uncanny valley" [63]. Digital animators are therefore presented with a paradox: aesthetically, highly illusionistic characters are better able to express nuance, but are more challenging to devise in ways which are not perceived as "creepy," whereas more schematic characters have the virtue of expressive clarity and audience acceptance through their status as obviously animated fictions. O'Neill suggests that "animated characters with a high degree of abstraction or a cartoon-like appearance are generally more accepted by the audience" because they avoid being insufficiently "lifelike [70:10].

The danger of such simplifications is evident in visual animations and cartoons, where the conjectural inferences often rely on caricatures: emotional expression, with its emphasis on the physicality of the body, is particularly prone to stereotyping in ways

that reinforce existing gendered or racialized hierarchies [67,68]. Generative AI systems for images clearly exhibit this bias [104]. This problem is also evident in textual animations like ChatGPT. Salvaggio [79] describes the outputs of generative AI systems for images like DALL-E as "infographic[s] about the dataset" they are generated from, which in turn "tell us stories about that dataset and the human decisions behind it." Text outputs from chatbots like ChatGPT are analogous: textual representations of the LLM's "grammar of action." Without content moderation, these systems also produce discriminatory, inflammatory, or bigoted text to the extent their training data is also discriminatory [73].

All "generative" AI is in fact inferential AI: a particular generative model is deploying a "grammar of action" based on probabilistic reasoning to conjecture the "best" output based on the user's input. Who is able both to make and generate conjectures and put those conjectures into forms that can classify and define others, has never been a more urgent question. Animations powered by generative AI systems thus risk accelerating reification of existing stereotypes [56,61]. If the same inferences are both being drawn about populations and then producing representations of those populations, then individuals will be under strong cultural and social pressure to conform to these particular "grammars of action."

3.3 Introjection by Design

While animated characters can be designed with the aim of eliciting particular responses from an audience, the spectator's projection of liveliness onto the character is often determinative. Animated objects tap into the human tendency to anthropomorphize or assign human qualities to inanimate objects. Such characters are powerful precisely because they provide a schematic, simplified form onto which an individual viewer can transfer their attention and project their own emotions. Silvio notes animation entails a high degree of human psychological projection: she draws on Winnicott's idea of transitional objects and psychic transference to suggest that animations function as "psychically projected objects of desire" for the viewer [88:427]. "Truly alive' versus 'the illusion of life' may ultimately be a meaningless distinction to the audience," declared Andrew Stern, one of the pioneers of developing emotive virtual agents in the early 2000s [100:353]. Humans are in any case expert at perceiving meaningful two-way communicative exchanges even when no interlocutor exists [18]. As noted above, highly representative digital animations of human beings may be less effective as objects of projection because they lack the clarity and simplicity of more schematized animated characters. However, the opposite appears to be true of text-based animation, with more sophisticated interactivity creating a greater illusion of sentience [58]. ChatGPT and similar systems are often enthralling, capturing a user's emotional and mental attention as the aforementioned exchange between Roose of the *Times* Microsoft's LLM-powered search bot illustrates almost too well.

Animated AI systems with a high degree of interactivity are often described in the language of anthropomorphization [1,83], and there is a growing literature on anthropomorphism and parasociality in digital agents, including chatbots. As Maeda and Quan-Haase observe, "anthropomorphic features operate as social affordances [...] that simulate reciprocal engagements and foster a sense of trust between users and chatbots" [57:2]. Parasocial or one-sided relationships between an individual and an animated character are more potent if the animated character is interactive: the responsiveness of even a simple chatbot such as ELIZA was notably compelling, and length of time the Sydney chatbot caught and held Kevin Roose's attention and emotional engagement [84] was central to the latter's sense of disquiet and fascination. Silvio suggests that "a psychic theory of animation focuses on the projection of the self into the environment," but when animated characters are interactive, they introduce an element of synthetic introjection into their communication with a human viewer [88:426–427]. Introjection is more characteristic of interpersonal reciprocity: but in the case of interactive animated agents, only the human actor is reflexively changing their perspective based on the actions of the animated entity. As such, the chain of influence is only one-way, from the creators of the animated character through it to the responses and self-conceptualization of the audience. Indeed, interactive AI agents could be understood as a form of attention trap akin to contemporary content recommender systems [84].

Discourses of anthropomorphization do implicitly acknowledge the degree of projection often expressed by human users in the context of digital systems in general and interactive agents in particular [93]. However, less critical attention has been paid to the impacts of designing interactive virtual agents of all kinds who are designed to draw out and heighten emotional projection in various digital contexts. Instead, a misguided focus on AI sentience has tended to confuse human projective capacity with AI intelligence. Silvio observes that, "in practice, performance and animation are [hard] to separate" [88:432]. When this conflation is integrated into interactive AI system as a design goal, the implications are wide-ranging and potentially disruptive [52]. Understanding interactive AI technologies like ChatGPT as forms of animation highlights the politics of their deployment and use [2,22,111] as agents intended to elicit a strong introjective emotional response. In particular, these systems have the potential to be exploited in the service of labor deskilling in the service sector, facilitate emotional manipulation in political and commercial contexts [15], and be used as vectors of disinformation [48].

4 GOVERNING ANIMATED AI SYSTEMS

Proposals for the governance of artificial intelligence technologies have proliferated in recent years, developed by national and subnational governments, international organizations, civil society groups, sectoral professional associations, and the business community [97,101,106,114]. Since the release of ChatGPT in late 2022, this discourse has been largely focused on the notion of "AI safety," whose proponents often presume the imminent development of Artificial General Intelligence (AGI) and the necessity of so-called "values alignment" between soon-to-be sentient AI systems and the interests of human stakeholders [13,81,96]. Recognizing technologies such as ChatGPT as animated entities or characters reframes the question of values alignment in practical way. Instead of centering hypothetical and self-aggrandizing concerns regarding the speculative impacts of future AGI, identifying interactive AI systems as animated entities places these technologies within the ambit of a variety of existing legal, regulatory, policy, and

design tools, including extant literature in human-computer interaction [18,38,44,57,60] and scholarship on values in digital design [3,23,26,27,28]. Moreover, acknowledging the impacts these interactive systems are having in the present underscores the urgency of crafting appropriate civic responses for the benefit of publics around the world.

If policymakers do accept ChatGPT and similar chatbots powered by Large Language Models (LLMs) are evocative and expensive textual animations, what avenues do they have for the governance and regulation of these and other interactive AI systems? Existing legal and regulatory frameworks seem likely to be necessary but not sufficient [40]. Rules around copyright and intellectual property law, product liability, and consumer labeling may all help delimit the appropriate capacities and uses of such AI technologies. For instance, laws in jurisdictions like California which already require chatbots to identify as such when interacting with the public [7], and recent FTC rulings have made the application of the agency's mandate to LLMs and AI more broadly explicit through analyses of design and user experience features [51]. However, some potential regulation of interactive representation may run afoul of democratic commitments to freedom of expression. One way to differentiate between appropriate and inappropriate animated AI systems may therefore the extent to which an animation is found to be unfair or deceptive in the context of consumer protection law [98]. Such a distinctions could also develop separate categories of evaluation for animations generated for passive consumption and those intended to interact dynamically with the public. Regardless, there is considerable scope for further scholarship on how to mobilize an array of existing and proposed governance strategies to ensure animated AI technologies are used appropriately and in common interest.

Given the particular persuasive power of various forms of anthropomorphic or quasi-anthropomorphic representation in animated AI systems, a commitment to "AI iconoclasm" is one possible, if speculative, governance principle. Iconoclasm is the act of breaking or destroying idols physically; the term also refers figuratively to the overthrow of received beliefs reference to beliefs or traditional institutions. Such an AI iconoclasm would question whether various forms of animated representation, such as sexually explicit animated deepfakes, are ever ethically or socially appropriate in the development of AI systems, and whether design measures intended to encourage the projection of human attributes onto an AI system might be deemed unethical in some contexts. It is not clear if even limited AI iconoclasm would be a practical measure within the current legal framework of entities such as the United States or the European Union; it is also not clear how such AI iconoclasm would be effect across different media modalities (including speech, image, or text generation). Nonetheless, the idea of AI iconoclasm has at least literary precedent [43], and is one possible avenue to catalyze a broader discussion regarding the social effects, positive and negative, of animated AI systems.

5 CONCLUSION

ChatGPT and other LLMs are evocative animations, but like all forms of animation, they present only the illusion of vitality. Claiming these technologies deserve recognition as persons makes as much sense as doing the same for a Disney film. We must therefore disenthrall ourselves and recognize the extent to which developments in artificial intelligence have long been, in fact, developments in increasingly persuasive multimodal animations. More research is needed to distinguish where and how various different types and examples of AI systems sit on the continuum between performance and animation, through analyses of their features, functionalities, and mechanisms [57,88] . The complexity of interactive AI systems suggests the need for multiple nuanced accounts of how these animated technologies support interactivity. However, by cutting through the AGI hype and recognizing precisely what these technologies are and how their capacities for engagement are limited, we can move forward with reality-based conversations about their social impacts: how they are best used, and how best to restrict their abuse in meaningful ways.

ACKNOWLEDGMENTS

This work has developed through and benefited from considerable feedback and constructive criticism, including from Annette Zimmerman, Benjamin Fish, Lisa Henderson, Pinar Barlas, Daniel Arauz Nuñez, Takuya Maeda, Tarleton Gillespie, Michael Anderson, and Joanna Redden; the fourth-year students in my Fall 2023 seminar on "The Ethics and Social Impacts of AI Systems" at Western University; from participants in the Rotman Institute for Philosophy 2023 Annual Library Lecture Series on Ethical Issues in Artificial Intelligence, the 2023 University of Guelph Centre for Advancing Responsible & Ethical Artificial Intelligence (CARE-AI) lecture series, and the Leibniz Center for Informatics-Schloss Dagstuhl Seminar (24121), "Trustworthiness and Responsibility in AI - Causality, Learning, and Verification" in March of 2024; and funding support from the Social Sciences and Humanities Research Council of Canada (SSSHRC) (grant # 430-2021-01041), the Canadian Institute for Advanced Research (CIFAR), the Schwartz Reisman Centre for Technology and Society at the University of Toronto, and the Faculty of Information and Media Studies (FIMS) and the Rotman Institute for Philosophy at Western University.

REFERENCES

[1.] Gavin Abercrombie, Amanda Cercas Curry, Tanvi Dinkar, and Zeerak Talat. 2023. Mirages: On Anthropomorphism in Dialogue Systems. *arXiv*. https://arxiv.org/abs/2305.09800

[2.] Philip E. Agre. 1994. Surveillance and capture: Two models of privacy. *The Information Society* 10, 2: 101–127.

[3.] Philip E. Agre. 1997. Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In G.C. Bowker, L. Gasser, S.L. Star, and B. Turner, eds., *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work*. Lawrence Erlbaum, 131–158.

[4.] Philip E. Agre. 1997. *Computation and human experience*. Cambridge University Press.

[5.] Anne Allison. 2006. *Millennial Monsters: Japanese Toys and the Global Imagination*. University of California Press, Los Angeles, and Berkeley.

[6.] Louise Amoore. 2022. Machine learning political orders. *Review of International Studies*: 1–17.

[7.] Robert Bateman. 2023. How to Comply with California's Bot Disclosure Law. *TermsFeed*. Retrieved May 1, 2024 from https://www.termsfeed.com/blog/ca-bot-disclosure-law/.

[8.] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*: 610–623.

[9.] Abeba Birhane, Jelle van Dijk, and Frank Pasquale. 2024. Debunking Robot Rights Metaphysically, Ethically, and Legally. *arXiv*. https://arxiv.org/abs/2404.10072

[10.] Ian Bogost. 2007. *Persuasive Games: The Expressive Power of Video Games*. The MIT Press.

[11.] Adam I. Bostic. 1998. Automata: Seeing Cyborg through the Eyes of Popular Culture, Computer-Generated Imagery, and Contemporary Theory. *Leonardo* 31, 5: 357–361.

[12.] Karen Brown. 2021. Something Bothering You? Tell It to Woebot. *The New York Times*. Retrieved from https://www.nytimes.com/2021/06/01/health/artificial-intelligence-therapy-woebot.html.

[13.] Brian Christian. 2020. *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton & Company, Inc., New York.

[14.] Simon Coghlan, Kobi Leins, Susie Sheldrick, Marc Cheong, Piers Gooding, and Simon D'Alfonso. 2023. To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digital Health* 9: 20552076231183544.

[15.] Vincent Conitzer, John Tasioulas, Matthias Scheutz, *et al.* 2022. Mimetic Models. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*: 479–490.

[16.] Alison Darcy, Jade Daniels, David Salinger, Paul Wicks, and Athena Robinson. 2021. Evidence of Human-Level Bonds Established with a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. *JMIR Formative Research* 5, 5: e27868.

[17.] Gilles Deleuze. 1990. Postscript on Control Societies. In *Negotiations, 1972-1990*. Columbia University Press, 177–182.

[18.] Jayson L. Dibble and Sarah F. Rosaen. 2011. Parasocial interaction as more than friendship: Evidence for parasocial interactions with disliked media figures. *Journal of Media Psychology: Theories, Methods, and Applications* 23, 3: 122–132.

[19.] Benj Edwards. 2023. Warning of AI's danger, pioneer Geoffrey Hinton quits Google to speak freely. *Ars Technica*. Retrieved May 1, 2024 from https://arstechnica.com/information-technology/2023/05/warning-of-ais-danger-pioneer-geoffrey-hinton-quits-google-to-speak-freely/.

[20.] Lelia Erscoi, Annelies Véronique Kleinherenbrink, and Olivia Guest. 2023. Pygmalion Displacement: When Humanising AI Dehumanises Women. https://osf.io/preprints/socarxiv/jqxb6

[21.] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health* 4, 2: e19-19.

[22.] Mary Flanagan, Daniel C Howe, and Helen Nissenbaum. 2008. Embodying Values in Technology: Theory and Practice. In J. van den Hoeven and J. Weckert, eds., *Information Technology and Moral Philosophy*. Cambridge University Press, 322–353.

[23.] Mary Flanagan and Helen Nissenbaum. 2014. *Values at Play in Digital Games*. The MIT Press.

[24.] Juliet Floyd. 2017. *Philosophical Explorations of the Legacy of Alan Turing*. Springer, 1–51.

[25.] Adam Frank, Andrew Stern, and Ben Resner. 1997. *Socially Intelligent Virtual Petz*. AAAI Technical Report FS-97-02.

[26.] Batya Friedman and David G. Hendry. 2019. *Value Sensitive Design*. The MIT Press.

[27.] Batya Friedman and Helen Nissenbaum. 1996. Bias in Computer Systems. *ACM Transactions on Information Systems* 14, 3: 330–347.

[28.] Batya Friedman and Helen Nissenbaum. 1997. Software Agents and User Autonomy. *AGENTS '97: Proceedings of the first international conference on Autonomous agents*, 466–469.

[29.] Dina Genkina. 2024. AI Prompt Engineering is Dead. *IEEE Spectrum*. Retrieved from https://spectrum.ieee.org/prompt-engineering-is-dead.

[30.] Judith Genova. 1994. Turing's sexual guessing game. *Social Epistemology* 8, 4: 313–326.

[31.] Ilana Gershon. 2010. *The Breakup 2.0*. Cornell University Press.

[32.] Ilana Gershon. 2015. What Do We Talk about When We Talk About Animation. *Social Media + Society* 1, 1: 1–2.

[33.] Carlo Ginzburg. 2009. Morelli, Freud, and Sherlock Holmes: Clues and Scientific Method. *History Workshop Journal* 9, 1: 5–36.

[34.] Ben Goertzel. 2014. Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence* 5, 1: 1–48.

[35.] Trystan S Goetze. 2024. AI Art is Theft: Labour, Extraction, and Exploitation, Or, On the Dangers of Stochastic Pollocks. *arXiv*. https://arxiv.org/abs/2401.06178

[36.] John-Stewart Gordon and Ausrine Pasvenskiene. 2021. Human rights for robots? A literature review. *AI and Ethics*: 1–13.

[37.] Nico Grant and Cade Metz. 2022. Google Sidelines Engineer Who Claims Its A.I. Is Sentient. *The New York Times*. Retrieved from https://www.nytimes.com/2022/06/12/technology/google-chatbot-ai-blake-lemoine.html.

[38.] Jonathan Gratch and Stacy Marsella. 2004. A domain-independent framework for modeling emotion. *Cognitive Systems Research* 5, 4: 269–306.

[39.] Michael M. Grynbaum and Ryan Mac. 2023. The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work. *The New York Times*. Retrieved from https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

[40.] Philipp Hacker, Andreas Engel, and Marco Mauer. 2023. Regulating ChatGPT and Other Large Generative AI Models. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 1112–1123.

[41.] Kali Hays. 2023. Andreessen Horowitz would like everyone to stop talking about AI's copyright issues, please. *Business Insider*. Retrieved January 18, 2024 from https://www.businessinsider.com/marc-andreessen-horowitz-ai-copyright-2023-11.

[42.] Will Douglas Heaven. 2023. The inside story of how ChatGPT was built from the people who made it. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/.

[43.] Brian Herbert and Kevin J. Anderson. 2002. *Dune: The Butlerian Jihad*. Tor Books, New York.

[44.] Katherine Isbister. 2006. *Better Game Characters by Design*. Elsevier/Morgan Kaufmann Publishers.

[45.] Margot E. Kaminski, Matthew Rueben, William Smart, and Cindy Grimm. 2017. Averting Robot Eyes. *Maryland Law Review* 76: 983–1025.

[46.] William Keith. 1994. Artificial intelligences, feminist and otherwise. *Social Epistemology* 8, 4: 333–340.

[47.] Amy Kind. 2022. Computing Machinery and Sexual Difference: The Sexed Presuppositions Underlying the Turing Test. In K. Maitra and J. McWeeny, eds., *Feminist Philosophy of Mind*. Oxford University Press, Oxford, and New York.

[48.] David Klepper. 2023. It turns out that ChatGPT is really good at creating online propaganda. *Fortune*. Retrieved from https://fortune.com/2023/01/24/chatgpt-open-ai-online-propaganda/.

[49.] Will Knight. 2023. Some Glimpse AGI in ChatGPT. Others Call It a Mirage. *Wired*. Retrieved from https://www.wired.com/story/chatgpt-agi-intelligence/.

[50.] Michal Kosinski. 2023. Theory of Mind May Have Spontaneously Emerged in Large Language Models. *arXiv*. https://arxiv.org/abs/2302.02083

[51.] Shreyansh Kotak. 2023. Navigating the FTC's new rules on deceptive patterns in subscription cancelations. *Medium*. Retrieved May 1, 2024 from https://uxdesign.cc/navigating-the-ftcs-new-rules-on-deceptive-patterns-in-subscription-cancelations-fc48d0c4a256.

[52.] P M Krafft and Joan Donovan. 2020. Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. *Political Communication* 37, 2: 194–214.

[53.] Brenda Leong and Evan Selinger. 2019. Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism. *Proceedings of the Association for Computing Machinery's Conference on Fairness, Accountability, and Transparency* (2019): 299-308

[54.] Lydia H Liu. 2011. *The Freudian Robot: Digital Media and the Future of the Unconscious*. The University of Chicago Press.

[55.] Samantha Lock. 2022. What is AI chatbot phenomenon ChatGPT and could it replace humans? *The Guardian*. Retrieved from https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans.

[56.] Donald A. MacKenzie. 2016. Performing Theory? In *An Engine, Not a Camera*, The MIT Press, 1–35.

[57.] Takuya Maeda and Anabel Quan-Haase. 2024. When Human-AI Interactions Become Parasocial: Agency and Anthropomorphism in Affective Design. *2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*.

[58.] Kyle Mahowald, Anna A Ivanova, Idan A Blank, Nancy Kanwisher, Joshua B Tenenbaum, and Evelina Fedorenko. 2023. Dissociating language and thought in large language models: a cognitive perspective. *arXiv*. https://arxiv.org/abs/2301.06627

[59.] Dan Mangan. 2024. Microsoft, OpenAI sued for copyright infringement by nonfiction book authors in class action claim. *CNBC News*. Retrieved January 15, 2024 from https://www.cnbc.com/2024/01/05/microsoft-openai-sued-over-copyright-infringement-by-authors.html?__source$=$sharebar

[60.] Paul Manning and Ilana Gershon. 2013. Animating Interaction. *HAU: Journal of Ethnographic Theory* 3, 3: 107–137.

[61.] Leo Marx. 2010. Technology: The Emergence of a Hazardous Concept. *Technology and Culture* 51, 3: 561–577.

[62.] Brian Merchant. 2024. The AI industry has a battle-tested plan to keep using our content without paying for it. *The Los Angeles Times*. Retrieved from https://www.latimes.com/business/technology/story/2024-01-12/column-copyright-is-the-biggest-threat-to-the-ai-industry-but-its-not-going-down-without-a-fight.

[63.] Masahiro Mori. 2012. The Uncanny Valley. *IEEE Spectrum*. Retrieved from https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley.

[64.] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are Social Actors. *Human Factors in Computing Systems*: 72–78.

[65.] Mark Neocleous. 2003. The Political Economy of the Dead: Marx's Vampires. *History of Political Thought* 24, 4: 668–684.

[66.] Katherine Newey and Jeffrey Richards. 2010. *John Ruskin and the Victorian Theatre*. Springer.

[67.] Sianne Ngai. 2002. "A Foul Lump Started Making Promises in My Voice": Race, Affect, and the Animated Subject. *American Literature* 74, 3: 571–602.

[68.] Sianne Ngai. 2005. *Ugly Feelings*. Harvard University Press.

[69.] Helen Nissenbaum. 2011. From Preemption to Circumvention. *Berkeley Technology Law Journal* 26, 3: 1367–1386.

[70.] Rob O'Neill. 2008. *Digital Character Development: Theory and Practice*. Morgan Kaufmann/Elsevier.

[71.] OpenAI. 2023. GPT-4 System Card. https://cdn.openai.com/papers/gpt-4-system-card.pdf.

[72.] Will Oremus. 2022. Google's AI passed a famous test — and showed how the test is broken. *Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2022/06/17/google-ai-lamda-turing-test/.

[73.] Billy Perrigo. 2023. OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. *Time*. Retrieved from https://time.com/6247678/openai-chatgpt-kenya-workers/.

[74.] Frank Proschan. 1983. The Semiotic Study of Puppets, Masks and Performing Objects. *Semiotica* 1–4, 47: 3–44.

[75.] Byron Reeves and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Center for the Study of Language and Inference.

[76.] Jennifer Rhee. 2018. *The Robotic Imaginary:The Human and the Price of Dehumanized Labor*. University of Minnesota Press, Minneapolis, MN.

[77.] Jessica Riskin. 2016. *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*. University of Chicago Press, Chicago.

[78.] Kevin Roose. 2023. Bing's Chatbot Drew Me In and Creeped Me Out. *The New York Times*, A1. Retrieved from https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html.

[79.] Eryk Salvaggio. 2022. How to Read an AI Image: The Datafication of a Kiss. *Cybernetic Forests*. Retrieved January 9, 2024 from https://cyberneticforests.substack.com/p/how-to-read-an-ai-image.

[80.] Michael Schade. How ChatGPT and Our Language Models Are Developed. *OpenAI*. Retrieved January 21, 2024 from https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed.

[81.] Bruce Schneier and Nathan Sanders. 2023. The A.I. Wars Have Three Factions, and They All Crave Power. *The New York Times*. Retrieved May 1, 2024 from https://www.nytimes.com/2023/09/28/opinion/ai-safety-ethics-effective.html.

[82.] Maximilian Schreiner. 2023. GPT-4 architecture, datasets, costs and more leaked. *The Decoder*. Retrieved January 20, 2024 from https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/.

[83.] Gabriela Scorici, Mario D. Schultz, and Peter Seele. 2022. Anthropomorphization and beyond: conceptualizing humanwashing of AI-enabled machines. *AI & SOCIETY*: 1–7.

[84.] Nick Seaver. 2019. Captivating algorithms: Recommender systems as traps. *Journal of Material Culture* 24, 4: 421–436.

[85.] Terrence J. Sejnowski. 2023. Large Language Models and the Reverse Turing Test. *Neural Computation* 35, 3: 309–342.

[86.] Andrew D. Selbst, danah boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. *2019 ACM Conference on Fairness, Accountability, and Transparency (FAT '19)*, 59–68.

[87.] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. 2023. The Curse of Recursion: Training on Generated Data Makes Models Forget. *arXiv*. https://arxiv.org/abs/2305.17493

[88.] Teri Silvio. 2010. Animation: The New Performance? *Journal of Linguistic Anthropology* 20, 2: 422–438.

[89.] Teri Silvio. 2019. *Puppets, Gods, and Brands: Theorizing the Age of Animation from Taiwan*. University of Hawaii Press, Honolulu, HI.

[90.] Tom Sito. 2013. *Moving Innovation: A History of Computer Animation*. The MIT Press.

[91.] Luke Stark. 2018. Algorithmic psychometrics and the scalable subject. *Social Studies of Science* 48, 2: 204–231.

[92.] Luke Stark. 2018. Facial recognition, emotion, and race in animated social media. *First Monday* 23, 9.

[93.] Luke Stark. 2020. Here Come the Computer People: Anthropomorphosis, Command, and Control in Early Personal Computing. *IEEE Annals of the History of Computing* 42, 4: 53–70.

[94.] Luke Stark. 2023. ChatGPT is Mickey Mouse. *Daily Nous*. Retrieved May 1, 2024 from https://dailynous.com/2023/03/16/philosophers-on-next-generation-large-language-models/#stark.

[95.] Luke Stark. 2023. Artificial intelligence and the conjectural sciences. *BJHS Themes* 8: 35–49.

[96.] Luke Stark. 2023. Breaking Up (with) AI Ethics. *American Literature* 95, 2: 365–379.

[97.] Luke Stark, Daniel Greene, and Anna Lauren Hoffmann. 2021. Critical Perspectives on Governance Mechanisms for AI/ML Systems. In J. Roberge and F. McKelvey, eds., *The Cultural Life of Machine Learning*. Palgrave Macmillan, 257–280.

[98.] Luke Stark and Jevan Hutson. 2022. Physiognomic Artificial Intelligence. *Fordham Intellectual Property, Media, and Entertainment Law Journal* 32, 4: 922–978.

[99.] Julia Steiner. 2024. Fair Use Friction in the Age of Artificial Intelligence: The New York Times v. OpenAI. *University of Miami Law Review Blog*. https://lawreview.law.miami.edu/fair-use-friction-in-the-age-of-artificial-intelligence-the-new-york-times-v-openai/

[100.] Andrew Stern. 2003. Creating Emotional Relationships with Virtual Characters. In R. Trappl, P. Petta, and S. Payr, eds., *Emotions in Humans and Artifacts*. The MIT Press, Cambridge, MA, 333–362.

[101.] Kristen Thomasen. 2023. Safety in Artificial Intelligence & Robotics Governance in Canada. *SSRN Electronic Journal*.

[102.] E.R. Truitt. 2016. *Medieval Robots: Mechanism, Magic, Nature, and Art*. University of Pennsylvania Press, Philadelphia.

[103.] Alan M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX, 236: 433–460.

[104.] Victoria Turk. 2023. How AI reduces the world to stereotypes. *Rest of World*. Retrieved January 2, 2024 from https://restofworld.org/2023/ai-image-stereotypes/.

[105.] Sherry Turkle. 2005. *The Second Self: Computers and the Human Spirit*. The MIT Press.

[106.] Lena Ulbricht and Karen Yeung. 2021. Algorithmic regulation: A maturing concept for investigating regulation of and through algorithms. *Regulation & Governance*.

[107.] Evelyn Wan. 2021. "I'm like a wise little person": Notes on the Metal Performance of Woebot the Mental Health Chatbot. *Theatre Journal* 73, 3: E-21-E-30.

[108.] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9, 1: 36–45.

[109.] Joseph Weizenbaum. 1976. *Computer Power and Human Reason*. W.H. Freeman and Company.

[110.] Elizabeth A. Wilson. 2010. *Affect and Artificial Intelligence*. University of Washington Press, Seattle, WA.

[111.] Langdon Winner. 1988. Do Artifacts Have Politics? In *The Whale and the Reactor*. Chicago, Il, 19–39.

[112.] Thomas Woodside and Helen Toner. 2024. *How Developers Steer Language Model Outputs: Large Language Models Explained, Part 2*. Center for Security and Emerging Technology, Georgetown University, Washington, DC.

[113.] Hannah Zeavin. 2021. *The Distance Cure: A History of Teletherapy*. The MIT Press, Cambridge, MA.

[114.] Anneke Zuiderwijk, Yu-Che Chen, and Fadi Salem. 2021. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly* 38, 3: 101577.