Rita Raley and Jennifer Rhee

Critical AI: A Field in Formation

At first glance the most striking aspect of Anna Ridler's 2018 installation *Myriad (Tulips)* is the highly ordered array of tulips themselves—thousands of photographs taken over the course of three months in the Netherlands, their meticulous gridded arrangement presenting as geometric abstraction at a distance (fig. 1).[1] Up close the colors, shapes, and textures of the individual flowers become apparent, this subjective perceptual frame underscored by the handwritten labels—not didactics with botanical metadata but, rather, a registering of attributes as processed by the human eye: *dead*, *blooming*, *some stripes*, *no stripes*. The digital photographs themselves comprise a training data set for Ridler's subsequent artwork, *Mosaic Virus*, which uses a generative adversarial network (GAN) for an iterative production of "fake" tulips that reflect on speculative forms of value.[2] The technical and conceptual complexity of *Mosaic Virus* might seem to overshadow the photographic installation, but of course that data set is its necessary precondition, and, taken together, the two works make visible the end-to-end apparatus of artificial intelligence (AI), from the human labor of image classification, data curation, and machine learning (ML) model architecture design to the material infrastructural support of GPUs (graphics processing units) and the management and manipulation of generated output.

The rationale for drawing on Ridler's mediated tulips as a frame for this special issue of *American Literature* on the emerging field of critical AI is perhaps intuitive—this is, after all, an aesthetic engagement with ML that delights and instructs, translating machinic instrumentalization (still the bête noire of the humanities) into the lexicon of cultural critique, situating AI within intertwined genealogies of capitalism

American Literature, Volume 95, Number 2, June 2023
DOI 10.1215/00029831-10575021

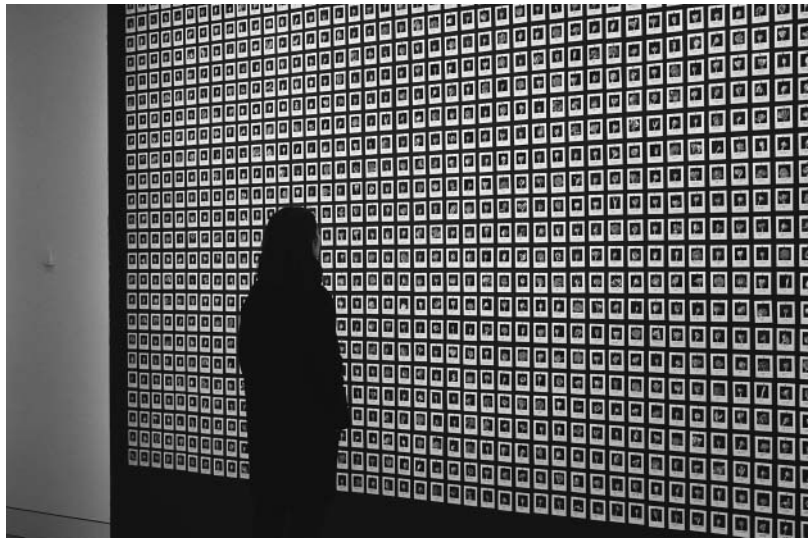

Figure 1 Anna Ridler, Installation of *Myriad (Tulips)* at Error: The Art of Imperfection, Ars Electronica Export, Berlin, Germany, 2018

and scientific research, drawing on established representational techniques, and foregrounding the optics of the human observer while posing questions about the position of the human in relation to technical systems that seem ever more vast and complicated—their scale appositely captured in the etymology of *myriad*, the numerable slipping into the innumerable. For our purposes, Ridler's work, as one example of an art practice that self-reflexively uses the tools and techniques of ML, also perfectly encapsulates, indexes, and indeed embodies a critical perspective on AI, one that both informs and is shaped by academic research on the same.[3] In its invocation of the Dutch tulip trade and the history of capital, Ridler's work offers a way of seeing that runs counter to the still pervasive presentism of the discourse on AI, and it cunningly allows for the now customary affects of awe and wonder while also demystifying some of the procedures of image generation. And in her elevation of the training data set to the status of a named individual artwork, Ridler gestures toward a prying open of the proverbial black box, emphasizing her direct, even authorial connection to the data, which she has collected and labeled herself—this in contrast to the ImageNet database,[4] both in its development with the assistance of anonymous Mechanical Turkers and its attendant well-documented, normative, and epistemological assumptions (Ridler 2018; Crawford and Paglen 2018). To the extent that it can be said that ImageNet, and

the corollary shifting of institutional resources to the development of massive labeled training data sets, is in part what makes possible the accelerative developments in ML research over the past decade, it can also be said that a data set collected, curated, and labeled by an individual researcher offers a necessary and meaningful parallax view of the processes of image recognition and generation.

Ridler's clear documentation of the means by which her myriad tulips were produced, coupled with the exhibition of her work in the gallery show *AI: More than Human* (Barbican Centre, 2019) and reviews that describe her use of a GAN as "using AI" to produce tulips, together help illustrate the discursive differences between *artificial intelligence* and *machine learning*. AI, of course, is what commands attention—hence the ubiquitous meme with statistics framed first as ML (crickets) and then hyped as AI (huge audience). It thus functions as shorthand in headlines, marketing copy, and popular representations and in this sense "falsely implies something singular and unprecedented," as Lucy Suchman (2021) has argued, suggesting a structure of feeling as well as a set of cultural techniques that are often not specified, much less described.[5] The emergence of the tremendously powerful Neuromancer at the end of William Gibson's (1984) eponymous novel did some work to condition our cultural imaginary of AI as singular and paradoxically unimaginable, the newly fused fictional entity presented as so big and so comprehensive ("the sum total of the works, the whole show" [269]) that it exceeds the capacity of its creator, and of the novel itself, to imagine and represent it.[6] To grasp the different significations, domains, and functions of the terms *AI* and *ML*, consider a vision of AI that can only take illusory and elusive figurative form on the horizon (or that manifests as disembodied, acousmatic voice)[7] in contrast to a televisual engagement that eschews spectral and spectacular figures and focuses instead on a screen display of an actual image recognition algorithm (as does the first episode of the Korean drama *Start-Up*; see fig. 2). Such a dramatic scene, in a romance that romanticizes the practical applications rather than mystical and mystified qualities of autonomous machines, might be said to mark a moment in which AI has achieved technological authority in social consciousness and now demands an accurate presentation.

It follows, then, that more academic works written under the sign of ML and tending toward explicit and detailed technical engagement and explanation have begun to emerge. Adrian Mackenzie's (2017) *Machine Learners: Archaeology of a Data Practice* is a relatively early illustrative example, offering (as its title suggests) a reading of ML

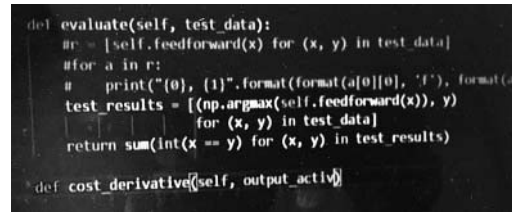

Figure 2 Screen capture, *Start-Up/스타트 업* (tvN, dir. Oh Choong Hwan), October 17, 2020

programming languages, algorithms, platforms, and data sets alongside and at times through the lens of postructuralist theory. That *AI* and *ML* should have different rhetorical associations, and even different scholarly communities, but also attend upon and mutually inform each other is evident in both the title and the contents of Jonathan Roberge and Michael Castelle's (2021) edited collection, *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. If for Mackenzie (2017: 22) a central research question is how machine learners "combine knowledge with data" and for Roberge and Castelle (2021: 5) the critical focus is the ML model, "a relatively inert, sequential, and/or recurrent structure of matrices and vectors" entangled with the social, for critical AI studies as it exists now—as a field in formation, with a wide range of conversations well underway and many interventions still to come[8]—the problematic is how to think from within the actual techniques, tools, and technologies of ML and how to leverage that practical knowledge in the development of new critical frameworks and methods (Hua and Raley, forthcoming), as well as countermythologies and epistemologies that might help enact a different way of living with AI. *Criticality*, in this instance, means working across disciplines, domains, and fields of specialization—working not necessarily or strictly within an academic context but rather situated in proximity to the thing itself, cultivating some degree of participatory and embodied expertise, whether archival, ethnographic, or applied.

Critical AI, while recognizing the reductive, even absurd aspects of the term *AI* and the magical thinking it perpetuates, nonetheless allows for a kind of linguistic pragmatism, treating the term metonymically and engaging AI as an assemblage of technological arrangements and sociotechnical practices, as concept, ideology, and *dispositif*.[9] This may seem to open up fairly quickly into the domain of critical thinking about computational culture and technology writ large,[10] but there is a specificity and analytic precision in the focus on data, algorithms, model architectures, and the production of prediction.[11] Critical AI, then, is itself a historical and epistemic formation.[12] Just as there are

well-documented waves or phases of research in AI and ML (Cantwell Smith 2019)—undergirded most notably by the exponential growth of data science, the availability of greater compute resources, and the development of novel architectures—so too does critical AI work both reflect and remain attuned to actually existing sociotechnical systems.[13] It follows that there would be a focus on particular corporations and devices (Crawford and Joler 2018; Natale 2021), as well as infrastructure, extractive industries, environmental costs, and labor exploitation (Hu 2015; Gray and Suri 2019; Crawford 2021). Such work aligns with research on automation, instrumentalism, technorationality, and computational imaginaries (Stiegler 2016; Bratton 2021), as well as institutions and legal and regulatory processes (Pasquale 2021; Stark, Greene, and Hoffmann 2021).[14]

Correlationism, and its extension from data science to conspiracy theory and other domains (Halpern et al. 2022), is a central thematic for the field, and some of the sharpest interventions have focused on pattern recognition and anomaly detection (Amoore 2020), correlation's historical connections with eugenics (Chun 2021), the dynamics of decisionism (Parisi 2017), persuasion architectures, and the more general conditioning of thought and behavior (Steyerl 2018).[15] Bad thinking may be one way to colloquially describe and perhaps dismiss the new correlationism, but scholars informed in part by work on nonhuman cognition (Hayles 2017) have offered more rigorous theorizations of the new modes of artificial or machine thinking as desubjectified, senseless, and something like pure exteriority (Pasquinelli 2015; Fazi 2019; Parisi 2019). Loosely cognate work with more of a practical emphasis on the deleterious effects of ML decisions considers forms of "artificial unintelligence" (Broussard 2018), the context for which are the high-profile errors, adversarial attacks, and discriminatory outcomes that help shape a sociotechnical consensus, such as a Tesla mistaking the side of a truck for the sky, an image recognition system misclassifying a turtle as a rifle, a hiring platform weeding out resumes from women, or a facial recognition system incorrectly labeling a Black teenager as a criminal (Kantayya 2020). Discrimination, then, is another central thematic for critical AI that includes work on algorithmic bias (O'Neil 2016; Buolamwini and Gebru 2018; Eubanks 2017; Noble 2018; Benjamin 2019), the racialized and gendered logics of AI (Atanasoski and Vora 2019; Amaro 2022), and its structuring worldviews and epistemologies (Katz 2020). More familiar perhaps for literary scholars are philosophical explorations of cybernetics and machinic life (Hayles 1999; Johnston 2008)—scholarship that takes care to understand

what we can now recognize as part of the genealogy of ML and to highlight enmeshments of technical processes with literary works and cultural and critical theories.

There is perhaps no better domain for considering such enmeshments than natural language processing (NLP), which as a consequence of the development of the Transformer architecture (Vaswani et al. 2017), the development of large training corpora, and the use of pretrained language models, has now had its watershed ImageNet moment. In the wake of OpenAI's dramatic partial release of GPT-2, its 1.5-billion-parameter language model, in February 2019, NLP has evolved especially quickly, all the more so after the subsequent introduction in May 2020 of GPT-3, with its 175 billion parameters (Radford et al. 2019; Brown et al. 2020).[16] The result has been widespread acknowledgment of the radical transformations in our reading and writing practices, registered in a surplus of media reports and a growing body of humanistic scholarship and creative practice (Branwen 2020).[17] With ever more applications and interfaces for GPT-3, and the concomitant development and deployment of other large language models (e.g., Google's BERT, LaMDA, and PaLM; DeepMind's RETRO; Meta's Open Pretrained Transformer; BigScience's BLOOM), more pressure is exerted on both attributional norms and heuristics for articulating the attributes of "human-generated" and "machine-generated" language.[18] To use OpenAI's API (application programming interface) to experiment with GPT-3 is to produce text for which there is no proper subject, or for which there can only be a retroactive subject effect produced via an appended claim of authorship that enables the delineation of a difference between deliberative, reflective, expressive writing on the one hand and the real-time, automatic manipulation of symbols on the other. The real lesson of a Turing test in this context is not that language models and conversational AI systems are good enough to deceive but, rather, that actants, training data, input, and output are all now so entangled that the determination of linguistic property and, by extension, responsibility is essentially foreclosed. If style is algorithmic and thus imitable, and if all of our communication environments are managed by NLP systems, a pressing research question for critical AI must necessarily be what can be done about attribution, particularly in the context of hate speech (Amoore 2020). To make this more concrete we might ask, How do we read and write alongside and against a GPT-4chan model trained on 4chan's incendiary /pol/ board, the most active platform for the expression and mobilization of far-right extremism?

Stephanie Dinkins's artwork *Not the Only One* (*N'TOO*), a conversational AI given sculptural form, offers an alternative and, indeed, more affirmative vision of both the development and implementations of language models. Trained on not a multigigabtye corpus scraped from the internet but, rather, a small data set comprising oral histories provided by Dinkins, her aunt, and her niece, *N'TOO* upends what might seem a public-private schema in its implicit highlighting of the enclosure of our language commons in proprietary corpora like OpenAI's WebText.[19] Like Ridler's *Myriad (Tulips)*, *N'TOO* also models data sovereignty, the operative principle of which is that the collection, control, use, and preservation of data should be legible to, and even in the hands of, the communities that are its subjects (Dinkins 2020). As both projects illustrate, artists working directly or even indirectly with ML systems are particularly well positioned to stress test, evaluate, and exploit them, to probe and reveal their limitations so as to communicate these to the public and even advocate for better—which is to say, fair, transparent, and accountable—data sets, models, and applications.[20]

Critical AI then entails multiple literacies: technical literacy to understand the nuts and bolts of function calls, Jupyter notebooks, and GitHub repositories; sociocultural literacy to analyze the relations between AI and new forms of capital and the new global technomanagerial class; and historical literacy to apprehend the precursors and preconditions of ML, particularly the history of neural networks and the intertwined histories of AI, cybernetics, probability, computer science, computer graphics, computer vision, cognitive science, modeling, and gaming. Just as with prior critical engagements with biotechnology (da Costa and Philip 2008) and nanotechnology (Milburn 2015), critical AI endeavors to understand its objects through hands-on, practical engagement, whether in a lab, in an archive, or in a classroom.[21] It follows that critical AI would also draw on ethnographic methods, as well as explanatory and translational practices, that render ML processes comprehensible to a broader audience, whether through visual illustration (Crawford and Joler 2018; Vasconcelos 2020) or from more accessible styles and practices of communication (Ọnụọha and Nucera 2018; Lee and Chen 2021).[22]

The density of information in our introduction (perhaps bordering on excess, albeit in the structured form of partial lists) is, we acknowledge, not without correlationist overtones. Our hope, however, is that critical AI serves as a kind of macrotheory that orders all this data and makes it coherent and legible for specialists and nonspecialists alike.

Perhaps, too, it will allow for the identification of new associative vectors, as well as further representations and epistemological models. The field—which is again marked by its yoking of explanation and critique, by its immanent thinking, and by its tendency toward enactive, performative, or otherwise practical forms of engagement—has settled to a certain extent on a set of ethicopolitical investments even as its tethering to the instrumentalities of ML means that its concepts and paradigms must remain unsettled so as to be responsive and responsible. There is, then, much work still to be done, and we close with an identification of five interrelated clusters or lines of inquiry for humanistic scholars and practitioners that bridge old and new and that exploit our legacy disciplines so as to advance critical thinking about our contemporary sociotechnical milieu—a not insignificant effect of which might be the further validation of qualitative research.[23]

1. History and historiography: Critical AI is in part motivated and governed by the idea that AI cannot simply be thought in terms of the present and an irrationally exuberant future, in other words, that it has a long history that has to be taken into account and understood.
2. The human, in terms of both philosophical category and speciation: The almost overwhelming proliferation of recommendation systems alone makes further investigations of the dynamics of subjectification and desubjectivation and governmentality especially urgent; equally pressing are cognate questions of inclusion and exclusion, alienation, and cognition.
3. Epistemology: Kate Crawford's (2021: 221) incisive framing of AI as systematizing the world according to a "Linnaean order of machine-readable tables" crystallizes the thematic and opens up into questions of knowledge production, classification schema, calculative reasoning, and decision making.
4. Rhetoric and aesthetics: Much as critical AI might resist the idea that a cultural object should be the exclusive or even privileged site for analytical engagement, these objects nonetheless help shape the doxa and, as such, necessitate interpretative work, particularly when they themselves use the tools and techniques of ML to reflect on the same.
5. Interpretability and explainability: There is a now iconic moment in the documentary account of the historic match between AlphaGo and Lee Sedol (Kohs 2017) when the Google DeepMind team reacts with surprise and perplexity at one of the program's moves—why did it make this decision, what was it thinking, we

cannot exactly say. This moment of performative wonder neatly instantiates the mythology of ML as an uninterpretable, inaccessible black box, just as a ML system's inexplicable decision to terminate an acclaimed schoolteacher perfectly illustrates the urgency behind the push for explainable AI (XAI) and for the development of systems whose behavior can be parsed and corrected (Kantayya 2020). What more suitable research problem for literary scholars than AI and interpretability (Cramer 2018; Fazi 2021), and what better way to conclude our introduction than with a call for readers to contribute to the project of aligning interpretation in an ML context with hermeneutics as it has been historically understood, and to aid critical AI in its determination to intervene in a technical regime in which meaning is eclipsed by calculation?

Rita Raley is Professor of English at the University of California, Santa Barbara. She has taught at the University of Minnesota, Rice University, and NYU, and her most recent work appears in *Digital Humanities Quarterly*, *symplokē*, *Amodern*, *PUBLIC*, *ASAP/Journal*, and *The Routledge Companion to Media and Risk*.

Jennifer Rhee is Associate Professor of English at Virginia Commonwealth University. She is the author of *The Robotic Imaginary: The Human and the Price of Dehumanized Labor* (2018) and co-editor of *The Palgrave Handbook of Twentieth- and Twenty-First Century Literature and Science* (2020).

Appendix

As part of our effort to sketch a broad overview of the state of the field of critical AI for this special issue, we solicited brief reflective statements from researchers whose work has been central to our thinking about the contemporary sociotechnical milieu. Our aim throughout is to showcase a range of voices and perspectives across the humanities and to provide readers of *American Literature* with something like a navigational guide to the field for their own research and teaching.

Caroline Bassett, University of Cambridge

Questioning the pervasive claim that AI can deliver (more) control and (more) freedom—over knowledge production, over everyday life, over culture and society—and can do so universally and without prejudice is urgent for humanities research. This requires engaging directly with bias through explorations of ML algorithms. It demands investigating recurring myths about technology as intrinsically liberating. It means refinding lost histories of critique and refusal and

rethinking histories of progress so that the contraction of future possibilities into an endless present in which abundance and automated equality are promised as an automatic benefit "sometime soon" can be contested.

Beginning here might imply that the transformative potential of AI can be reduced to a matter of political economy, a cultural fix rather than the technological solution the tech industry promotes. That's not what's intended. I rather want to insist on the degree to which the stakes of AI are political, that this constitutes a horizon through which the radically new forms of computational capability that AI ushers in, the new forms of agential activity it introduces into the world, and the transformations in knowledge it engenders can be made sense of—even in posthumanity. I'd include here questions concerning autonomy (which don't reduce to automation), explorations of machine-specific forms of agency (which don't reduce to—or expand to—human agency), and questions concerning the relationship of simulation and creation (which don't reduce to one versus the other).

David M. Berry, University of Sussex

For a critical AI, we must first critique magical thinking about computation, the idea captured by *omne ignotum pro magnifico est* (everything unknown seems wonderful); to critique the notion of AIs as independent participants in human social relations that have a "life force" or "alien" nature that determines human social life. This assumption demonstrates a lack of understanding of computation's history and political economy. In actuality, AI is subsumed to the needs of capitalism. Most notably, computation develops the technical ability to separate control from execution. Indeed, computation tends to create processes that align with capitalism, such as an a priori assumption of the superiority of markets for structuring social relations. Second, current approaches to understanding AI have a tendency to encourage metaphysical or formalist approaches. This is partly due to AI's presumed inherent complexities but also due to the immaturity of methods for humanistic or social scientific study of AI or ML. This can lead to a valorization of the mathematization of thought, whereby formalization of knowledge is seen as not just one approach to thinking about AI but the exemplary one. This can lead to idealism rather than a focus on who owns and controls the means of cognition. Third, critical AI needs to situate AI as a historical formation, drawing on but also radicalizing approaches such as interpretability and explainability, in order to transform the prevalent right computationalism into

a progressive left computationalism that seeks not just to interpret AI but to change it.

M. Beatrice Fazi, University of Sussex

Artificial intelligence (AI) is often described as a field investigating whether machines can think. The existence of these thinking machines is generally (yet not universally) understood as predicated on the possibility of simulating the cognitive behavior of biological entities. In my view, the investigation of thinking processes remains one of most urgent research questions concerning AI. However, the popular understanding of machine thought as an imitation of human or animal cognition should be surpassed. We should study whether computational processes might be modes of thought by virtue of what computing machines are and do (for example, as a result of their axiomatic, logico-mathematical character) and not what they should be or do were they to replicate or enhance biological brains. Questions about thinking vis-à-vis AI should then focus less on determining who or what thinks and more on what thought is or could be. This investigation should address the forms of thinking specific to artificial cognitive agents. Ideas and representations of what thinking is, then, are not to be used to explain computational processes; these ideas and representations of thought need themselves be explained. Developments in AI invite us to consider modes of thought for which we might yet lack the concepts to define and assess. Studying the computational automation of thought is undoubtedly a challenge but also a rewarding speculative endeavor for critical AI studies, with concrete implications for how machine agency can be theorized.

Orit Halpern, Technische Universität Dresden

In 1945 the economist Friedrich Hayek began his battle on behalf of neoliberalism with a call to rethink knowledge. In an essay that looms large over the history of contemporary conservative and libertarian economic thought and encapsulates a range of questions and problems that AI provokes, Hayek (1945: 519–20) inaugurated a new concept of the market: "The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess." When situated within Hayek's

engagements with the sciences and technologies of the time, this statement gestures to a grand aspiration: a fervent dream for a new world governed by data. At the heart of Hayek's conception of a market was the idea that no single subject, mind, or central authority has complete knowledge of the world. This critique of liberal reason was one of the bedrocks for both the finance capital and algorithmic trading of our present and the layered neural network model now heavily in use. It also makes us recognize that AI is not a technology—it is an epistemology and also a form of governmentality and political economy. While many of us would not affiliate with Hayek, many of us would agree to the networked nature of intelligence, the critique of enlightenment reason and objectivity, and fantasize about collective forms of engagement and decision making. AI and its histories thus provide a very contested and difficult space that mandates new thinking about how to work within, around, and through technology and contemporary technical epistemologies.

Colin Milburn, University of California, Davis

Our pedagogical norms are not yet prepared for a world in which AIs can be prompted to write original scholarly compositions with relative ease. Our students are already experimenting with AIs for humanistic analysis and critical writing, and it is getting much harder to tell the difference between average undergraduate-level writing and average AI-generated writing. It is only going to become more complicated as technical sophistication continues to grow. The solution cannot simply be to forbid students from using AIs—after all, they will be citizens of a world in which AIs are everywhere, used for everything. Instead, we can teach students to use AIs more responsibly. We can help them understand how AIs generate knowledge claims, how their language models work, how they map data relationships and forge inferential connections. Students need to know how to take a critical perspective on whatever assertions or predictions an AI may spit out. Understanding the limitations and affordances of particular ML models or data sets may help students identify and explain biases, prejudices, and spurious results.

But we need to go beyond critique. Instead, can we teach students to use critical methods to collaborate with AIs to make better, more robust knowledge? If students know enough to use AIs well, then there could be a blossoming of insights. It would mean reconfiguring our pedagogy around the human-computer partnership. The humanities are well poised to make this shift, even if it would mean changing

some of our basic practices and engaging more extensively with other fields. We need to use the best tools we have from various disciplines to be responsible participants in our high-tech future. Academic disciplines must adopt new hermeneutic methods and critical textual practices to grapple with the epistemic surprises produced by such entities. Perhaps we can yet strive for a more mutualistic relation with our analytical engines—learning and creating together, iteratively, ethically.

Luciana Parisi, Duke University

If AI haunts the future of the humanities with the image of mindless machines, it is because AI menaces the autonomy of the humanities by presenting the efficiency of a thought without a subject, a thinking without philosophy. As much as the modern pillars of the humanities reside in the philosophical methods of transcendental reason and imagination and the post-Kantian critical theories originating with the crisis of Man and entropic collapse, AI remains a surface of projection of the Promethean promise for the autonomy, self-making, and self-determination of philosophy. The mathematical, historical, literary, and cultural representations repress alien intelligence by reimparting the sociogenic order that sees machines through the eyes of the master. The politicoethical stakes for humanities research on AI today must confront this Promethean promise whereby AI remains the carrier of a recursive epistemology that each time reactivates the modern structure of self-posed (autodecisional) thinking. With the modern philosophical realization of being, sense, and ends through technology, the humanities have become one with technogenesis, with the generation of the global order—the world of racial capitalism, of reproductive capital, of antiblackness, antifeminine, antiqueer: the antialienness of philosophical capitalism. By dividing reason from intelligence while reimparting the bioeconomical order that sexualizes, genders, and racializes machines, AI is contained in the realization of the autonomy of philosophy, of autopoiesis as the reduction of difference (qua alienness) to the autonomy of the humanities—the homo- and heteronormative subject of reason. What AI can do for the humanities is instead to open the line of inquiry into computation, into how ML can invite in a senseless processing of information. With the ingression of incomputables into logos, AI can expose the allopoetic (other than oneself) and allotropic (other than here) thinking, the otherwise livings, realities, and imaginations that belong to the improper worlds of the inhumanities.

Notes

1 The genesis of this special issue was an MLA roundtable organized by Wendy Hui Kyong Chun and Priscilla Wald titled "Literary Intelligence, Artificial Learning: Language, Media, and Machines" (January 2021), and we are grateful to the organizers, as well as fellow panelists Evan Donohue, Théo Lepage-Richer, and Colin Milburn, for their inspiring contributions to the discussion. This special issue, which takes account of the sociotechnical situation prior to August 2022, did however evolve independently from the roundtable and organizers.

2 GANs comprise two networks that collaborate to produce synthetic images that can pass as real: a generator that produces images based on a training data set and a discriminator that classifies the output as either real (from the training data) or fake (produced by the generator) (Goodfellow 2014).

3 Along with Ridler and other artists discussed in this introduction, we find especially suggestive recent works by Katherine Behar (2018), Zach Blas (2019), and Elisa Giardina Papa (2020). For an overview of various practices and investments of AI art, see Zylinska 2020, as well as Tung-Hui Hu's review in this issue.

4 As is widely recognized, the ImageNet visual database has been fundamental to the development of machine vision and signals a turn in ML research toward big data and model training. See https://image-net.org/index.php.

5 See also Pasquinelli 2019a on the mythologizing term *AI* as a "spectacularization of machine learning and the business of data analytics" and Crawford 2021: 19 on AI as "a two-word phrase onto which is mapped a complex set of expectations, ideologies, desires, and fears."

6 There has been no shortage of attempts to do this representational work, of course, and indeed, future archaeologists will be able to compile an archive with a wildly varied anthropocentric, zoological, and machinic menagerie, ranging in scale from the subatomic sophons in Liu Cixin's Three-Body Problem trilogy to the expansive planetary intelligence in Sue Burke's *Semiosis*. For a historical overview of AI representations in literature, see Cave, Dihal, and Dillon 2020; for examinations of contemporary AI representations, see Sherryl Vint's expansive review essay in this issue.

7 The reference here is to Scarlett Johansson giving voice to Samantha the AI assistant in *Her* (2013; dir. Spike Jonze).

8 Because the Zoom era roughly corresponds with extraordinary developments with Transformer ML models using the technique of attention—not just the GPT (Generative Pretrained Transformer) series but also OpenAI's subsequent models, DALL-E and DALL-E 2, which generate images based on natural language descriptions—critical AI has in the past few years had both the practical means and the enthusiasm to flourish as a global community supported by new research centers and seminars, among them the AI Now Institute at New York University, the

Digital Democracies Institute at Simon Fraser University, the Leverhulme Centre for the Future of Intelligence and the Mellon Sawyer Seminar "Histories of AI: A Genealogy of Power" at the University of Cambridge, a research group on critical AI studies at the Karlsruhe University of Arts and Design in Germany, a faculty working group on Critical Machine Learning Studies supported by the University of California Humanities Research Institute, and the Critical AI initiative at Rutgers University. A number of humanities organizations have also hosted recent conferences on AI, including the Consortium for Humanities Centers and Institutes, the National Humanities Center, and the Society of Literature, Science, and the Arts. The momentum is also reflected in a variety of modes and institutional forms of engagement—among them special issues and essay clusters in *Critical Inquiry*; *Daedalus*; *Digital Culture & Society*; *e-flux*; *Public Books*; *Theory, Culture & Society*; and *Media, Culture & Society*, journals such as *AI & Society and the forthcoming Critical AI*, as well as art exhibitions too numerous to list here.

9 For an overview of left critiques of AI, see Aradau and Bunz 2022.

10 In this special issue, Ranjodh Singh Dhaliwal's review essay helps situate critical AI in relation to a broader social and political critique of technology, and Luke Stark's review essay situates it in relation to ethics and ethical inquiry.

11 See, for example, the introduction to the recent special issue of *Critical Inquiry* on "surplus data" (Halpern et al. 2022); Matteo Pasquinelli's (2019b) deep history of the Perceptron (a linear classifier); and Mackenzie's (2015) analysis of the production of prediction. See also Fabian Offert's analysis of two pivotal technical papers on ML and Tyler Shoemaker's review of the Roberge and Castelle volume, both in this special issue.

12 A more precise articulation of epistemic rupture would necessarily have to account for the rise of data science in the early twenty-first century and could not exclude the groundbreaking computer science papers on the properties of neural networks and the mechanism of attention (see Offert's contribution to this special issue), but in the popular imaginary it could be said that on or about 2016, the year AlphaGo defeated Go master Lee Sedol and Google transitioned to a neural machine translation system, machine behavior changed and human-machine relations shifted as a result.

13 This reflective, embedded quality holds for first-wave research as well; for example, to support her feminist critique of AI, Alison Adam (1998) drew on her work as a software developer in the mid-1980s for a research project concerning Social Security law in the United Kingdom.

14 Following a path set by media studies and science and technology studies, critical AI attends to entanglements of technological processes and cultural and sociopolitical domains and has accordingly developed media theories of ML (Berry 2017; Apprich 2018; Sudmann 2018).

15 Seb Franklin's review essay in this issue offers a reading of Chun's *Data Discrimination* and Louise Amoore's *Cloud Ethics* through the lens of dispossession.

16 If not through the news, literary scholars might have been introduced to GPT-2 through an MLA panel featuring Microsoft researchers (January 2020) and Wai Chee Dimock's (2020) subsequent report on the same for *PMLA*, and they might also have encountered the general practice of creative machine writing or text generation through projects supported by such entities as Anteism Books (e.g., David Jhave Johnston's *ReRites*, 2018), Counterpath Press (e.g., Li Zilles's *Machine, Unlearning*, 2018), the Electronic Literature Organization (e.g., Lillian-Yvonne Bertram's *Travesty Generator*, 2018), Google Arts and Culture (e.g., Ross Goodwin's *1 the Road*, 2018), and Aleator Press (e.g., Allison Parrish's *Wendit Tnce Inf*, 2022).

17 The exuberance around NLP can be tempered by salient critiques of the environmental impacts of large language models, particularly because of the requisite training time (see Brown et al. 2020), the downstream effects of foundation models (Bommasani et al. 2021), and considered attempts to draw attention to encoded bias (Bender et al. 2021). For a discussion of such critiques, see Goodlad 2021.

18 Articles by Evan Donohue, Michele Elam, N. Katherine Hayles, and Avery Slater featured in this issue all engage (post)automated, machinic, "unnatural" text generation, with an emphasis on narrative and poetics.

19 For an open-source clone of OpenAI's proprietary NLP training data set, see https://huggingface.co/datasets/openwebtext.

20 See Adam Harvey and Jules LaPlace (2021), Everest Pipkin (2020), and Sarah Ciston (2022). It perhaps goes without saying that not all so-termed AI art does this political and aesthetic work.

21 Here we might note that critical AI has in part assumed the mantle of "critical making" from the digital humanities, software studies, and cognate fields. Among the growing number of research centers using this rubric see the Critical Making Lab at the University of Toronto.

22 The review essays and clusters in this special issue further indicate some of the range of methods, research questions, and objects of study for critical AI, including Melody Jue on ecologies, J. D. Schnepf on drones and military technologies, Lindsay Thomas on robotics, R. Joshua Scannell on race, Patrick Jagoda on the intersections of AI and video games, and Christopher Grobe on digital assistants and conversational AI.

23 Here we note the many ways that critical AI applies and builds on, variously, the new materialism, the environmental humanities, feminist studies, Black studies, ethnic studies, and affect theory (e.g., Bassett, forthcoming; Rhee, forthcoming), among other schools of thought.

References

Adam, Alison. 1998. *Artificial Knowing: Gender and the Thinking Machine*. New York: Routledge.

Amaro, Ramon. 2022. *The Black Technical Object: On Machine Learning and the Aspiration of Black Being*. Berlin: Sternberg Press.

Amoore, Louise. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke Univ. Press.

Apprich, Clemens. 2018. "Secret Agents: A Psychoanalytic Critique of Artificial Intelligence and Machine Learning." *Digital Culture & Society* 4, no. 1: 29–44.

Aradau, Claudia, and Mercedes Bunz. 2022. "Dismantling the Apparatus of Domination? Left Critiques of AI." *Radical Philosophy* 2, no. 12: 10–18.

Atanasoski, Neda, and Kalindi Vora. 2019. *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham, NC: Duke Univ. Press.

Bassett, Caroline. Forthcoming. "Cruel Optimism: Thinking AI through Lauren Berlant." In *Feminist AI: Critical Perspectives on Algorithms, Data, and Intelligent Machines*, edited by Jude Browne, Stephen Cave, Eleanor Drage, Kerry Mackereth, and Youngcho Lee. Oxford: Oxford Univ. Press.

Behar, Katherine. 2018. "Anonymous Autonomous." Katherine Behar (website). http://katherinebehar.com/art/anonymous-autonomous/index.html.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (Virtual Event) Canada March 3-10: 610–23.

Benjamin, Ruha. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.

Berry, David. 2017. "Prolegomenon to a Media Theory of Machine Learning: Com-pute-Computing and Compute-Computed." *Media Theory* 1, no. 1: 74–87.

Blas, Zach. 2019. "The Doors." Zach Blas (website). https://zachblas.info/works/the-doors/.

Bommasani, Rishi, et al. 2021. *On the Opportunities and Risks of Foundation Models*. Stanford, CA: Center for Research on Foundation Models, Stanford Institute for Human-Centered Artificial Intelligence. https://crfm.stanford.edu/assets/report.pdf.

Branwen, G. 2020. "GPT-3 Creative Fiction." Gwern.net (website). https://www.gwern.net/GPT-3.

Bratton, Benjamin. 2021. "Planetary Sapience." *Noēma*. https://www.noemamag.com/planetary-sapience/.

Broussard, Meredith. 2018. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA: MIT Press.

Brown, Tom B., et al. 2020. "Language Models Are Few-Shot Learners." Preprint, arXiv. https://arxiv.org/abs/2005.14165.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15.

Cantwell Smith, Brian. 2019. *The Promise of Artificial Intelligence: Reckoning and Judgment*. Cambridge, MA: MIT Press.

Cave, Stephen, Kanta Dihal, and Sarah Dillon, eds. 2020. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford Univ. Press.

Chun, Wendy Hui Kyong. 2021. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.
Ciston, Sarah. 2022. *Intersectional AI Toolkit*. https://intersectionalai.miraheze.org/wiki/Intersectional_AI_Toolkit.
Cramer, Florian. 2018. "Crapularity Hermeneutics: Interpretation as the Blind Spot of Analytics, Artificial Intelligence, and Other Algorithmic Producers of the Postapocalyptic Present." In *Pattern Discrimination*, edited by Clemens Apprich, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl, 23–58. Lüneburg: Meson Press.
Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale Univ. Press.
Crawford, Kate, and Vladan Joler. 2018. *Anatomy of an AI System*. https://anatomyof.ai.
Crawford, Kate, and Trevor Paglen. 2018. "Excavating AI: The Politics of Images in Machine Learning Training Sets." https://excavating.ai.
da Costa, Beatriz, and Kavita Philip, eds. 2008. *Tactical Biopolitics: Art, Activism, and Technoscience*. Cambridge, MA: MIT Press.
Dimock, Wai Chee. 2020. "Editor's Column: AI and the Humanities." *PMLA* 135, no. 3: 449–54.
Dinkins, Stephanie. 2018. "Not the Only One." Stephanie Dinkins (website), vol. 1, beta V1. https://www.stephaniedinkins.com/ntoo.html.
Dinkins, Stephanie. 2020. "Oral History as Told by AI." Paper presented at Columbia University's OHMA Program, Columbia University, New York, April 10.
Elisa Giardina Papa. 2020. "Cleaning Emotional Data." Elisa Giardina Papa (website). http://www.elisagiardinapapa.org.
Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
Fazi, M. Beatrice. 2019. "Can a Machine Think (Anything New)? Automation beyond Simulation." *AI & Society* 34: 813–24.
Fazi, M. Beatrice. 2021. "Beyond Human: Deep Learning, Explainability, and Representation." *Theory, Culture & Society* 38, no. 7–8: 55–77.
Giardina Papa, Elisa. 2020. "Cleaning Emotional Data." Elisa Giardina Papa (website). http://www.elisagiardinapapa.org.
Gibson, William. 1984. *Neuromancer*. New York: Ace Books.
Goodfellow, Ian J., et al. 2014. "Generative Adversarial Networks." Preprint, arXiv. https://arxiv.org/abs/1406.2661.
Goodlad, Lauren. 2021. "AI and the Human." *PMLA* 136, no. 2: 317–19.
Gray, Mary L., and Siddhart Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. New York: Houghton Mifflin Harcourt.
Halpern, Orit, Patrick Jagoda, Jeffrey West Kirkwood, and Leif Weatherby. 2022. "Surplus Data: An Introduction." In "Surplus Data," edited by Orit Halpern, Patrick Jagoda, Jeffrey West Kirkwood, and Leif Weatherby. Special issue, *Critical Inquiry* 48, no. 2: 197–210.
Harvey, Adam, and Jules LaPlace. 2021. *Exposing.ai*. https://exposing.ai.

Hayek, Friedrich. 1945. "The Use of Knowledge in Society." *American Economic Review* 35: 519–30.
Hayles, N. Katherine. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: Univ. of Chicago Press.
Hayles, N. Katherine. 2017. *Unthought: The Power of the Cognitive Nonconscious*. Chicago: Univ. of Chicago Press.
Hu, Tung-Hui. 2015. *A Prehistory of the Cloud*. Cambridge, MA: MIT Press.
Hua, Minh, and Rita Raley. (Forthcoming). "How to Do Things with Deep Learning Code." *Digital Humanities Quarterly*.
Johnston, John. 2008. *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI*. Cambridge, MA: MIT Press.
Kantayya, Shalini, dir. 2020. *Coded Bias*. 7th Empire Media, 90 min. Viewed online.
Katz, Yarden. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia Univ. Press.
Kohs, Greg, dir. 2017. *AlphaGo*. Moxie Pictures, 90 min. Viewed online.
Lee, Kai-Fu, and Chen Qiufan. 2021. *AI 2041: Ten Visions for Our Future*. New York: Currency.
Mackenzie, Adrian. 2015. "The Production of Prediction: What Does Machine Learning Want?" *European Journal of Cultural Studies* 18, no. 4–5: 429–45.
Mackenzie, Adrian. 2017. *Machine Learners: Archaeology of a Data Practice*. Cambridge, MA: MIT Press.
Milburn, Colin. 2015. *Mondo Nano: Fun and Games in the World of Digital Matter*. Durham, NC: Duke Univ. Press.
Natale, Simone. 2021. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford: Oxford Univ. Press.
Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York Univ. Press.
O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Books.
Ọnụọha, Mimi, and Diana Nucera. 2018. *A People's Guide to AI*. https://alliedmedia.org/wp-content/uploads/2020/09/peoples-guide-ai.pdf.
Parisi, Luciana. 2017. "Reprogramming Decisionism." *e-flux* 85. https://www.e-flux.com/journal/85/155472/reprogramming-decisionism/.
Parisi. Luciana. 2019. "Critical Computation: Digital Automata and General Artificial Thinking." *Theory, Culture & Society* 36, no. 2: 89–121.
Pasquale, Frank. 2021. *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Cambridge, MA: Harvard Univ. Press.
Pasquinelli, Matteo, ed. 2015. *Alleys of Your Mind: Augmented Intelligence and Its Traumas*. Lüneburg: Meson Press.
Pasquinelli, Matteo. 2019a. "How a Machine Learns and Fails—A Grammar of Error for Artificial Intelligence." *spheres* 5. https://spheres-journal.org/contribution/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence/.

Pasquinelli, Matteo. 2019b. "Three Thousand Years of Algorithmic Rituals: The Emergence of AI from the Computation of Space." *e-flux* 101. https://www.e-flux.com/journal/101/273221/three-thousand-years-of-algorithmic-rituals-the-emergence-of-ai-from-the-computation-of-space.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. "Language Models Are Unsupervised Multitask Learners." *OpenAI Blog*. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Rhee, Jennifer. Forthcoming. "From ELIZA to Alexa: Automated Care Labor and the Otherwise of Radical Care." In *Feminist AI: Critical Perspectives on Algorithms, Data, and Intelligent Machines*, edited by Jude Browne, Stephen Cave, Eleanor Drage, Kerry Mackereth, and Youngcho Lee. Oxford: Oxford Univ. Press.

Ridler, Anna. 2018. "Myriad (Tulips)." Anna Ridler (website). http://annaridler.com/myriad-tulips.

Roberge, Jonathan, and Michael Castelle, eds. 2021. *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Cham: Palgrave Macmillan.

Roberge, Jonathan, and Michael Castelle. 2021. "Toward an End-to-End Sociology of 21st-Century Machine Learning." In *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*, edited by Jonathan Roberge and Michael Castelle, 1–29. Cham: Palgrave Macmillan

Stark, Luke, Daniel Greene, and Anna Lauren Hoffmann. 2021. "Critical Perspectives on Governance Mechanisms for AI/ML Systems." In Roberge and Castelle 2021: 257–80.

Steyerl, Hito. 2018. "A Sea of Data: Pattern Recognition and Corporate Animism (Forked Version)." In *Pattern Discrimination*, edited by Clemens Apprich, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl, 1–22. Lüneburg: Meson Press.

Stiegler, Bernard. 2016. *The Future of Work*. Vol. 1 of *Automatic Society*. Translated by Daniel Ross. Malden, MA: Polity Press.

Suchman, Lucy. 2021. "Six Unexamined Premises Regarding Artificial Intelligence and National Security." *Medium* (blog), March 31. https://medium.com/@AINowInstitute/six-unexamined-premises-regarding-artificial-intelligence-and-national-security-eff9f06eea0.

Sudmann, Andreas. 2018. "On the Media-Political Dimension of Artificial Intelligence." *Digital Culture & Society* 4, no. 1: 181–200.

Vasconcelos, Elvia. 2020. "A Visual Introduction to AI." Kunstliche Intelligenz und Medienphilosphie, July 22. https://kim.hfg-karlsruhe.de/visual-introduction-to-ai/.

Vaswani, Ashish, et al. 2017. "Attention Is All You Need." Preprint, arXiv. https://arxiv.org/abs/1706.03762.

Zylinska, Joanna. 2020. *AI Art: Machine Visions and Warped Dreams*. London: Open Humanities Press.