Michele Elam

Poetry Will Not Optimize, or What Is Literature to AI?

Keywords artificial intelligence, literature, art, race, GPT-3

Fiction as Friction

At first glance it seemed no different than any other MLA [Modern Language Association] session: in a midsize room at the Washington State Convention Center, well attended but not quite filled to capacity, with people leafing through their programs, checking their phones, drifting in and out. It was session 388, "Being Human, Seeming Human." Arranged by the Office of the Executive Director, it was the first of its kind. Four of the six speakers were from Microsoft, expressly invited to start a conversation about what it means for those who self-identify as human to share the planet with those who seem to be.

—Wai Chee Dimock, "Editor's Column: AI and the Humanities" (2020)

Wai Chee Dimock's timely essay in *PMLA* marks the significance of Microsoft representatives being invited to the annual convention of the Modern Language Association. Her essay is also a call to the MLA membership—over twenty-five thousand members in one hundred countries, primarily academic scholars, professors, and graduate students who study or teach language and literatures—inviting richer engagement with technologists, with exponential technologies, and with the outsize impact of tech on nearly every aspect of private and public life.

American Literature, Volume 95, Number 2, June 2023 DOI 10.1215/00029831-10575077 © 2023 by Duke University Press

At the time of the MLA convention, the company OpenAI had only recently launched GPT-2 (abbreviation for Generative Transformative Pretrainer), cutting-edge technology that generates human-like text, which was at the center of the session's conversation. Yet only a few months after Dimock's *PMLA* essay was published, OpenAI released an even more advanced predictive language modeling system, GPT-3, putatively one hundred times more powerful than its predecessor. Even as of the writing of this article, GPT-3 has already been superseded by PaLM as well as by new models enabling text-to-image such as DALL-E 2 (see Ramesh 2021), Stable Diffusion, Midjourney, and HuggingFace and by Google's Imagen Video, generative AI that vields high-quality text-to-video.² All this gives fresh immediacy and urgency to cultural conversations about the significance of artificial intelligence (AI) for the arts and humanities, especially given the everexpanding universe of visual art, performance, music, symphonies, playscripts, film scripts and all genres of literature generated and augmented by AI.

As Stephen Marche (2021b) put it in "The Computers Are Getting Better at Writing," these extremely powerful innovations in language processing, changing the sociotechnical landscape, are nothing less than "vertiginous" and, whatever else we may think of it, should not be underestimated as some kind of a "toy." Even as we debate what AI-generated and -augmented literature is/is not/might be, it cannot be dismissed as simply a trending subfield, novel genre, or specialized interest; nor does it fall easily within the category of digital humanities. The ubiquity of socially transformative technologies' engagement with humanities has made the subject nearly unavoidable.

That vast sphere of influence is impacting how we think about language itself. Marche's characterization of how scientists think AI is changing the way we relate to literature and the role of the writer points to what some see as a kind of rhetorical colonization: "GPT-3 shows that literary style is an algorithm" and understands the role of the writer as "an editor almost . . . executing on your taste. Not as much the low-level work of pumping out word by word by word."⁴ As Amita Gupta, a founder of Sudowrite, which uses GPT-3, describes it to Marche (2021b), "The artist wants to do something with language. The machines will enact it. The intention will be the art, the craft of language an afterthought." At a minimum, this approach to art—shifting its value from the apparently lowly craft of writing to intent—will strike many as reductive if not insulting, as will the ambitious conclusion that GPT-3 or its progeny might eventually function as a Romantic muse: "The oldest poems in the Western tradition, the Iliad

and the Odyssey, begin with an invocation to the muse, a plea for a mysterious, unfathomable other to enter the artist, taking over, conjuring language. GPT-3 is a mysterious, unfathomable other, taking over, conjuring language" (Marche 2021b)

Unsurprisingly, then, there is often a very particular, visceral reaction to GPT-3's most recent aspirations to literature—no matter that it comes with even its creator's self-deprecating claims to literary insufficiency—because the technology goes well beyond forays with natural language processing and text production.⁵ Some people are physiologically repelled by language generation that can increasingly seem at times convincingly indistinguishable from human production. They experience what Mashiro Mori called the *uncanny valley*, a queasiness that comes when a technological creation too closely approximates reality—at least what an individual takes as the boundary conditions for the real, when simulacra and sui generis appear to lose distinction.

Here I suggest essential challenges posed for AI by the arts (writ large to include literary, visual, performative, theatrical, graphic, musical) and, in turn, how AI might productively challenge the arts. Just as AI has invited debates about what constitutes or performs intelligences far beyond the Turing test, AI revives foundational questions in the arts and humanities about what is or is not literature or art; who or what can make it; how is it credentialed; how compensated; who arbitrates taste, value, valuation, proprietary content, and provenance (especially in the case of AI-generated art); who gets to decide the arbitrators; and who (or what) counts as a maker. These are not abstract questions, and the stakes are high.

Let me offer a couple examples of how the arts challenge AI. First, many have pointed out that storytelling is always needed to make meaning out of data, and that is why humanistic inquiry and AI are necessarily wed. Yet, as N. Katherine Hayles (2021: 1605) writes, interdependent though they may be, database and narrative are "different species, like bird and water buffalo." One of the reasons, she notes, is the distinguishing example of indeterminacy. Narratives "gesture toward the inexplicable, the unspeakable, the ineffable" and embrace the ambiguity, while "databases find it difficult to tolerate" (1605). As she explains, indeterminate data "that are not known or that elude the boundaries of pre-established categories—must be either represented through a null function or not be represented at all"; data relies on "enumeration, requiring explicit articulation of attributes and data values" (1605). This intolerance for indeterminacy, or *noise* as it is called, when it comes to ambiguity has serious implications for categorizing and

representing social identities such as race, ethnicity, or gender that challenge the enterprise of categorization itself.⁷

Literature especially challenges several of the assumptions informing AI development in some very specific ways that can offer humanistic complementarity. Consider, for instance, that literature does not aspire to a seamless user experience. In fact, it turns our attention to those seams we are seduced into not seeing. After all, fiction is not frictionless; poetry will not optimize.8 Humanities and arts value the thoughtful pause, not the push for speed and maximization; they encourage reflection over regulation (not to say, reflection precludes regulation); they tend to prioritize improvisation over pattern recognition, possibility over prediction, social good over capital gain, the acknowledgment of narrative perspective(s) versus tech's implied omniscient anonymity, what Alice Adams calls the "view from nowhere."9 It explores the complexities of individual choice over socalled personalization, ¹⁰ in which "knowing thyself" does not equal the "quantified self." 11 Literary achievement is indifferent to the mindset of efficiency and the "blessings of scale." 12

Indeed, optimizing and scaling are so often taken uncritically as the means and ends to success—in product pitches, they often acquire an incantatory quality lending them an almost unimpeachable authority in certain academic and tech industry circles—not just in AI technology but in corporate world building. Historically, literature has been not just indifferent but justifiably cynical of these kinds of approaches to doing and thinking. There is a long history of novels and short stories presciently critiquing the value system of speed, scale, maximization, and improving human performance dating well before the nineteenth century.¹³

The Literary Consequences of Algorithmic Ahistoricity

There already exists a vast universe of GPT engagements with literature—I have done a few myself. Gwern Branwen and Shawn Presser (2019) offer some of the earliest experimenting on their website. They include a "tutorial of retraining OpenAl's GPT-2 (a text-generating Transformer neural network) on large poetry corpuses to generate high-quality English verse." Their early ambitious effort involved Branwen "retraining GPT-2–117M on a Project Gutenberg corpus with improved formatting, and combined it with contemporary poem dataset based on Poetry Foundation website." It quickly became clear that with this broader, curated body of literature to train the

algorithms, zero-shot/few-shot experiments (translation: the algorithm learns almost immediately what was being asked of it) could yield AI-generated fiction, nonfiction, poetry, operas, music, and more in pretty much any known genre.

Branwen's many essays and wide-ranging how-to demos are welcome by many, especially because they offer specific fine-tuning technical advice to problem-solve both with GPT-2 and with GPT-3. Branwen and Presser (2019) note the much broader cultural implications for the latter, since GPT-3 capitalizes on what is known as raw, unsupervised data—it is a model, they argue, that can metalearn and thus putatively offer "an understanding of the world, humans, natural language, and reasoning."

In terms of literature, however, a limitation in this method is its assumption about what high-quality verse actually is or how it can be attained. For example, Branwen suggests the blessings of scale enabled by large foundational models can solve questions of both aesthetic value and verisimilitude with an ability to "approach humanlevel poems" (Branwen and Presser 2019) Many have already posed compelling critiques of scaling and large foundational models like GPT-3, especially with regard to its amplifications of bias and hate speech. He but even those approaches, as with Branwen's, leave entirely untouched and tacit the problematic assumptions about the Turing test of human mimesis as the standard by which to assess artistry. 15

This mimetic model is commonplace yet increasingly being challenged. Those working in fields related to AI, particularly cognitive psychology and neuroscience, frequently evoke that model when they refer to neural networks, in which programmers attempt to mirror (what they understand of) the brain's activity. Often one hears not that computers might benefit from mirroring or replicating the brain's processes but, rather, that the brain itself is a machine, or at least we ought to behave and bend toward it like one. This seems the case in Branwen's (Branswen and Presser 2019) invitation to understand creative action primarily in terms of technological input. 16 The human exchange with the interface becomes simply a mode of technoinstrumentalism in which writing a poem is a matter of submitting "prompts as programming," as Branwen (2022a) puts it. There is a certain devaluation of the durational labor involved in the creative act implied by such an attitude, illustrated by Amit Gupta's claim—quoted earlier—about GPT-3 enabling writers to bypass the "low-level work of pumping out word by word" (Marche 2021b), and echoed by Emad Mostaque, CEO of Stable Diffusion, when he announced that "So much of the world is creatively constipated, and we're going to

make it so that they can poop rainbows" by expediting the creative process (Qtd. In Roose 2022).

Related to this increasingly influential technological framing of the world is what I refer to as algorithmic ahistoricity, which has an outsize and, I think, particularly concerning consequence for literary sensibility and creation. Literature tends to resist representing history as the static, self-explanatory, sequential data points that are grist for predictive algorithms. Novelists, not to mention contemporary professional historians, usually understand history not as an inexorable teleological march forward in time and progress but as palimpsest, a Mobius strip, an ongoing and dynamic negotiation between pasts and presents. That stands in important contrast to AI's training sets, which are dehistoricized in particular ways. To be clear, of course one can train an algorithm on historically accurate data—that is not my point. Rather, the challenge lies with what gets counted as usable data in the first place: the historical information for training sets is necessarily treated as a set of static points—information already reduced and rendered interpretable as usable data. One can add new or different data but data itself are treated as ahistorical for the purposes of programming. It may seem counterintuitive to suggest that ahistoricity lies beneath algorithms used to predict a future, but this is, in effect, what occurs when an algorithm informed by data—or, specifically, the "tokens" comprising data—predicts and generates, for instance, judicial sentences, or bank loan eligibility, or an anticipated cluster of words. As Marche (2021a) cogently describes it:

The tool applied to solve many natural language processing problems is called a transformer, which uses techniques called positioning and self-attention to achieve linguistic miracles. Every token (a term for a quantum of language, think of it as a "word," or "letters," if you're old-fashioned) is affixed a value, which establishes its position in a sequence. The positioning allows for "self-attention"—the machine learns not just what a token is and where and when it is but how it relates to all the other tokens in a sequence. Any word has meaning only insofar as it relates to the position of every other word. Context registers as mathematics. This is the splitting of the linguistic atom.

So why and how does that nuclear effect really matter for literature? I argue that the results do not simply make for form without meaning—rather, it matters because it can *render meaning senseless*.

Let me offer an exemplar of why meaning made senseless is more than an annoyance: "Transformer AI poetry: poetry classics as reimagined and rewritten by an artificial intelligence," a GPT-3 adaptation of Maya Angelou's famous 1978 poem "Still I Rise" (Branwen 2022a). It takes the opening four lines as prompts and then generates texts that putatively extend it in style and content. The first section below is Angelou's original text as prompt; what follows in italics is generated by GPT-3:

"Still I Rise!"

You may write me down in history With your bitter, twisted lies, You may trod me in the very dirt But still, like dust, I'll rise. If you want me to know What you're puttin' me through, If you want me to share your agony, If you want me to be a part of it. All I need is the truth. When you know why I'm singin' About the hard times I've had, When you know why I'm singin' About ole Jim Crow When you really know why I'm singin' That will be the day. You are a black dam in the river of my freedom Too many people, you're towin'. I may not be around after a while, Because the river's runnin' deep. 'Cause I'm gonna ride that mutha Almighty river, And keep on singin' the dreamin' Of songs tryna change the reality

Despite GPT-3's predictive capacity for language generation, drawing on an enormous database to train its algorithm, GPT's "As I Rise!" collapses Black vernaculars and expressive forms across two centuries, flattening intergenerational significations, making mush of historical references and cultural meanings. The AI reduction of language to the passive voice of executability evacuates African American literary history in terms of form, content, interpretability, and, therefore, impact or influence in and on the world. It is drawing on a vast archive of African American literary form but turns it into a cringeworthy jumble of blues, Black power, racial uplift, and Ole Man River minstrel (not including Paul Robeson's subversive revisions of the song).¹⁷

This problem is perhaps most evident in the AI poem's dialectal representation of speech, not simply because it is insensible to historical iterations of African American Vernacular English. In fact, if that were the issue, a simple corrective would merely involve training the algorithm to disaggregate vernaculars by decades or region or other preferred filters. The real challenge, perhaps, is AI's inability to account for representation itself. Written dialectal speech, after all, is already thrice mediated: a representation of a representation of the spoken.¹⁸ Most important, the literary dialectal project—deciding how, if, and when to orthographically represent actual speech—indexes social more than sonic realities. For example, dialectal representation is not phonetic (which would be unreadable) but what linguists termed eve speech, since at least Chaucer's time historically signaling illiteracy or lower-class status. Zora Neale Hurston, among others, experimented with the form to free it from those associations in order to tap the rich cultural reservoir of linguistic communities.

But GPT-3 adds yet additional and different layers of mediation so that poetic verity—whatever truth telling the poem makes possible—is put at yet another remove. As James Baldwin ([1979] 1998: 782) put it, describing Black English, language indexes experience, and form takes the shape of its need: "A language comes into existence by means of brutal necessity, and the rules of the language are dictated by what the language must convey. . . . A people at the center of the Western world, and in the midst of so hostile a population, has not endured and transcended by means of what is patronizingly called a 'dialect.'"

To elaborate with another related illustration that I hope clarifies why algorithmic ahistoricity cannot be resolved by expanding a training set: Pulitzer-prize-winning playwright August Wilson is renowned for a series of ten plays representing Black life, each created for a different decade across the twentieth century. All his plays are evocatively, densely layered with vernaculars that capture the "rhythms, logic and linguistic structure of black speech" in order to "celebrate the poetry of everyday life," as one scholar explains (H. Elam 2006: 35). But despite Wilson's interest in capturing African American experience at certain historical moments, his characters' language is intentionally not rigidly specific to any particular time and place. In fact, the plays' metaphysics ground the action simultaneously in time and out of time. Representation's potency—whether literature, theater, performance, et cetera—functions in these liminal spaces. In this case, Wilson's poetics operate as a literary idiolect that is also linguistically

representative: "Even as Wilson records authentic black dialect and attends to historical detail, he employs patterns of language and rhythm that are particular to his dramaturgy. Phrases such as 'I ain't studying you,' he repeats from play to play. Thus a Wilson play requires actors who have the acumen for Wilson-speak and his specific formalism" (36). This "Wilson-speak" both enacts and signifies on the living transgenerational language systems that both bring into relief and reaffirm Black identities and cultures. 19

All to say, GPT-3's literary experiments have not "failed" because they do not meet some moving target of a literary standard, or because of technological insufficiency, but because GPT-3's approach to language can make it harder for people to name and navigate their realities. For Baldwin, Wilson, and many others, this is a question of what flourishes or not in the world, of what realities are possible or eclipsed, of what souls are seen or not.

The Real Real

All that said, on the subject of realities, AI can also serve as a bracing wake-up call to a settled status quo. In the professional world of art, it has upended business as usual, forcing some uncomfortable reckonings with the industry's core assumptions and canonized practices. For instance, there was much handwringing and gnashing of teeth in the professional art world over the sale for \$432,500 of *Portrait of Edmond Bellamy*, created by a GAN (generative adversarial network).²⁰ It sold for forty-five times over its estimate and made the esteemed auction house the first "to offer a work of art created by an algorithm" (Christie's 2018).

The sale revived perennial questions about art and aesthetics. In addition to those about authorship and cultural status, as mentioned at the outset of this article, AI continues to pose pressing questions about authenticity, provenance, value, and creator compensation. The business model in the art world is also being upended. As one recent article's title put it: "A.I. Has the Potential to Change the Art Business—Forever. Here's How It Could Revolutionize the Way We Buy, Sell, and See Art" (Schneider 2020). In fact, the piece, which explains seven ways in which AI can assist—from exhibition curation to value prediction—and is part of a larger *Artnet Intelligence Report* that includes a survey on the challenges of AI art authentication. There has also been development of international compensation standards for artists working in the digital realm to ensure equitable pay given the future of work in these



Figure 1 The Night Watch's missing panels, reconstructed with AI, are attached slightly adjacent to the original painting. Credit: Rijkmuseum/Reinier Gerritsen. https://www .rijksmuseum.nl/en/press/press-releases/for-the-first-time-in-300-years-the-night-watch-is -complete-again.

new mediums. That includes the increasingly popular use of blockchain technology and nonfungible tokens (NFTs) for the purchase of digital and so-called crypto arts, a financial vehicle "shaking up the art world" (Chow 2021).

But museums and galleries are still taking steps to prevent the particular kind of existential aversion that also has accompanied some AI-generated literary efforts, whether creative expressions or GANaugmented literary histories. Take, for example, the almost unanimous international acclaim of the recent use of AI to reconstruct one of Rembrandt's most renowned but disfigured paintings, Militia Company of District II under the Command of Captain Bannick Cocq (known commonly as *The Night Watch*). The announcement in spring 2021 that AI had been successfully deployed by the prestigious Rijkmuseum in Amsterdam, which owns the masterpiece, to recreate the damaged pieces in the style of Rembrandt, and that its chief scientist, Rob Erdmann, had personally trained the neural networks, reassured many in the arts world concerned about tampering with the piece's authenticity (fig. 1). Ironically, it was AI that was seen as preserving the real since, as Erdmann put it, normally they would have commissioned an artist to

recreate the missing pieces but "then we'd see the hand of the artist there. Instead, we wanted to see if we could do this without the hand of an artist. That meant turning to artificial intelligence" (Mattei 2021).

For some, this collaboration exemplifies how AI might augment and support the arts.²¹ But the lurking, unaddressed tension between where artistic authenticity begins and ends, and the potential threat of anyone confusing an original artwork—replete with what Walter Benjamin ([1936] 2008: 19–55) called the "aura" of the genuine and singular²²—with an AI reproduction, is signaled by the fact that the reconstructed pieces were hung next to but not allowed to touch the original and, following the exhibition, "will be taken down out of respect for the Old Master" (Matei 2021). As Erdmann said, "It already felt to me like it was quite bold to put these computer reconstructions next to Rembrandt" (quoted in Mattei 2021). Thus, even in this embrace of AI, it exists as a kind of third rail: its use still often requires the performance of deference to, and carefully monitored distinction from, the master.

Similarly, most authors using GPT-3 (in good faith at least, and as a nod to this anxiety over the real/fake) often signal the distinction using a different font and text size, making it clear which is their prompt and own writing and which is the AI text response. The original text, like the Rembrandt, is held at bay and at a tenuous remove lest we confuse the human and the AI.

Thus, the initial dismissal by many that GPT-3 does not remotely approximate literature, let alone intelligence, only belies the sense of a threat deferred (until inevitably some even more advanced technology emerges). Moreover, the too-quick dismissals that GPT fails the standards of literature or intelligence skirt the fact that both have never been self-evident givens. The moving definitional target for both literature and intelligence—or, more accurately, the fact that both are culturally negotiated phenomena used as shorthand for demonstrable standards, constituted and recruited for implicit purposes and particular ends of use, and necessarily mediated by evolving disciplinary mindsets—exacerbates the fact that they both have historically held uniquely powerful and problematic status as measures of humanity, of the "human." ²³ Certainly, AI-generated art, in particular, touches a social nerve because it taps into broader and legitimate anxieties about forgeries, deep fakes—connected to slippages between truth and lie that have vast political consequences. Shakespeare may have staged a play within a play to "catch the conscience" of a king, but the suspicion still lingers in some circles that performance is not a form of truth telling but, instead, deception.

292 American Literature

Yet I think the unease goes even deeper because, as mentioned, art has historically indexed humanity itself.²⁴ Since the Enlightenment, at least, poesy has been considered one of the highest, most complex forms of individual expression and cultural achievement. And precisely for this reason, poetry has been used as a measure of a person's (or a race's) humanity—or, in the case of African Americans and really anyone deemed nonwhite, of their less-than-human status. In other words, the stakes may go unnamed but are nonetheless high in the flurry of usually uneasy thought pieces on GTP's ability to generate form sans meaning, on language generation not literary work, on the nature of authorship (reprising with fresh anxiety the "death of the author"),²⁵ and on what sophisticated autodidactic neural networks, more generally, hold for the future of the humanities—and, to the degree that Dimock's piece identifies a shifting in the profession, for the future of work in the humanities,²⁶

Beyond "Doing as Saying"

Machine learning's conception and application of language are instrumentalist, unidirectional, executable: doing as saying. As Wendy Hui Kyong Chun (2006: 66) points out, "Unlike any other law or performative utterance, code almost always does what it says because it needs no human acknowledgement." The use imperative for why AI reduces language to code for human-computer interfaces clarifies the challenge, which is not with technologists' intent or AI's circumscribed approach to language but with the generalization of it as an implied explanation for how language and literature operate across all contexts. Also, the kind of autotelic closure that Hayles (2021: 1603) rightly points out is needed for much technological work is precisely what makes it nearly impervious to any critical understanding of either data's ontology or the in situ performative scenes of humancomputer interaction so crucial to understanding how AI databases are realized, recruited, and relevant. Rather, they simply present, fully formed and naturalized as factual, neutral descriptions of the world rather than its own world. The fact that its very particular world which comes with not just an embedded ontology but an epistemology, a way of knowing and experiencing—has increasingly come to stand in for the world writ large is what I think informs so much cultural anxiety about AI. As Ruha Benjamin has put it many times, it is as if we are being forced to live in the imagination of a very few.²⁷ It also offers some explanation for why the tech industry initially often

framed issues such as bias as either a discrete glitch to be fixed or an intractable social problem beyond the pale of technologists' ken or interest, a social issue revealed and handled by others once the technology is released into the wild.

The humanist concern is not handwringing over a fall from cultural power, although Hayles (2021: 1606) does note that some critics are concerned that "database will replace narrative to the extent narrative fades from the scene" as data, replacing classical Greek- and Romanera narrative's explanatory force in understanding world events, becomes essential in identifying large-scale phenomena. The problem is not with natural language prediction per se but with the increasing monopoly of that particularly structural approach to language systems. Partnered with corporate interests in pushing at scale particular kinds of intentionally "sticky," "addictive" storytelling, the content and the form of language increasingly lead to a culling of narratives and narrative forms that do not serve that addiction.

Certainly, expression forms flourish both on and outside these platforms. Hayles (2021: 1606) suggests that narratives of all kinds are high- and low-culture narratives so irrepressibly proliferate that they are "as ubiquitous in everyday culture as dust mites." But it would be hard to deny that unified (and unifying) industry-driven, mightily funded, financially incentivized storytelling—a powerful complex of profit imperatives and corporate marketing of unprecedented influence and reach—is dominating and narrowing of narrative options. It reflects a kind of singularity creep into language and literature.²⁸ I do not mean to be either cynical or presentist here. It is true that the nineteenth-century rise of mass culture generated a redundancy of a certain genre of narratives, particularly advancing plots that push rags-to-riches providential rise and American exceptionalism, so this potential narrowing of content is not new. This is part of AI's cultural genealogy, and it is one in which certain invested racialized, gendered narratives of modernity overwhelm others (see Elam 2022).

This genealogy is another reason that, in this age of AI, although there are more horizontal and representational forms of diversity—diverse platforms to view more diverse content created, produced, and represented by more diverse talent—the effect is not necessarily a leveling of power and opening of access. It is essential to acknowledge that those voices must still contend with long-embedded power structures and forces in place in academe, media, and the tech and entertainment industries, with and against which they must—and surely will—offer alternative and counter soundings, registers, and codings.

Coda: Beyond Optimization

One of these more hopeful soundings might include Vauhini Vara's 2021 article "I Didn't Know How to Write about My Sister's Death— So I Had AI Do It for Me," which generated a great deal of heated controversy over a GPT-3 experiment. It documents an attempt after years of struggle to help her put into words what her sister's passing meant to her. In an NPR interview, she discusses the initially unsatisfying experience—the predictive algorithmic program got stuck in a repetitive loop, for instance—but it also reflected back what she recognized as her own canned clichéd language that she had been offering as prompts, serving as an unflattering but revelatory mirror to her own prose. The algorithm also incorrectly generated language about her sister's life that was untrue, for instance, that she was an athlete. But, significantly, this only prompted Vara to think more about the process of truth telling itself. In short, the response to her prompts in turn prompted her; the AI-generated script did not edit but, rather, provoked to the extent she interpreted it as such. In occasionally illuminating ways, the process refracted back to her the limits and potential of how she had initially put her experience into print. She tapped Al's interactive possibilities in which the craft of writing was more than an "afterthought" (Marche 2021b); rather, she used it more as a conversant, interlocutor, de facto therapist. It learned as she trained it with input, but most important, it also, to her surprise, provided material that informed her about herself.

And as for what to make of the GPT-3 text? She says she edited it for length, for impact. But the last line was all GPT-3 (here in italics) and "she especially loved that last sentence because it contains so much" (Low 2021): "Once upon a time, my sister taught me to read. She taught me to wait for a mosquito to swell on my arm and then slap it and see the blood spurt out. She taught me to insult racists back. To swim. To pronounce English so I sounded less Indian. To shave my legs without cutting myself. To lie to our parents believably. To do math. To tell stories. *Once upon a time, she taught me to exist.*"

As Vara (2021) notes, although some of the GPT-3-produced language was uncannily akin to what could be produced by a human, even the ones that were not up to the standard of simulacra in fact had significance: sometimes in others' examples she read online, as well as in her experiment, the "language was weird, off-kilter—but often poetically so, almost truer than writing any human would produce." The literary value of the "almost truer" dimension appears in the last version of her own story, in which Vara reprints nine iterations of her vignette,

allowing readers to see the process and negotiation with AI that she and her editor use, such that the last version acquires meaning through repeat and revise—a kind of signifying on that which came before.²⁹

This particular practice of signification involving the inclusion of her own and the application's discarded drafts, precursors, iterations of, and variations on a published or performed piece is increasingly representative of what we might call the new genre of GPT-3 literary projects.³⁰ It suggests a renewed challenge to the notions of an originating moment in an artistic process, to the belief in a static final iteration that necessarily holds superior cultural or aesthetic status. Moreover, it helpfully pushes against the more tightly held idea(l), particularly cherished in the West, of "art" as only that which issues, sui generis, from a singular, taken-for-granted human author. 31 Vara's story illustrates when and where meaning making remains a durational, performative, collaborative process among author, audiences, contexts, and interpretative lenses. The idea that meaning is suspended in time, as if trapped in amber, residing fixed in authorial intent or encoded/entombed in text itself, was long ago debated and, for most scholars, settled as too limited an account of communication.

In that sense, at its best and perhaps most interesting, AI-generated literature and art might capitalize on how meaning is already and always an ongoing, mutually constitutive, interpretive event. In this case, at least, AI holds the possibility of becoming a generative interlocutor for the writer, enabling multivalent ways of communicating, in the higher interests of human play, insight, and creativity. Moreover, in service of those higher interests, the arts and humanities are essential in reframing the endless questions about just what intelligence or creativity is, about who, why, and what ends motivate those questions. If so, literature might help us understand the aims of AI beyond augmenting the human experience. AI, like all technologies, is a crucible of our world views, our social priorities, commitments, investments, and aspirations. As such, perhaps one of its greatest uses is to allow it to reflect us back to ourselves.

Michele Elam is the William Robertson Coe Professor of Humanities in the English department and faculty associate director of the Institute for Human-Centered AI at Stanford University. She has served as director of African and African American studies and director of the graduate Program in Modern Thought and Literature. Her books include Race, Work, and Desire in American Literature, 1860–1930 (2003), The Souls of Mixed Folk: Race, Politics, and Aesthetics in the New Millennium (2011), and The Cambridge Companion to James Baldwin (2015). Elam is currently completing a book project tentatively titled "Race-Making in the Age of AI."

Notes

- 1 As decolonial artist-technologist Amelia Winger-Bearskin describes it, cutting-edge technology is not always unequivocally a good; she calls it the "bleeding edge" of innovation (Mozilla Pulse, n.d.).
- 2 The language processing system, PaLM, short for Pathways Language Model, for instance, draws on neural networks trained on over 540 billion parameters compared with GPT-3's initial 175 billion (Narang and Chowdhery 2022).
- 3 I make a distinction, admittedly blunt if generally apt, between digital humanities and humanities' engagement with AI by drawing on Joanna Drucker's (2009: 6) conclusion that "digital humanities was formed by concessions to the exigencies of computational disciplines. Humanities played by the rules of computer science and formal knowledge."
- 4 Amit Gupta, one of the founders of Sudowrite, one of the programs using GPT-3, quoted in Marche 2021b.
- 5 Sam Altman, CEO of OpenAI, from the outset downplayed expectations for GPT-3, saying it has "serious weaknesses" and still makes "silly mistakes" (quoted from a tweet by Altman, reprinted in Deoras 2022). Nonetheless, the technology, which was licensed to Microsoft but invites (vetted) participation in the collective development of it, immediately gained cultural traction and immense popularity, especially among casual users.
- 6 The term *uncanny valley* was first coined by in the 1970s by Masahiro Mori, professor at the Tokyo Institute of Technology, who documented the physiological response of humans' affinity for social robots the more lifelike they appear—up to a point, after which affinity turns to repulsion. Since then, work has been extended and explored by social psychologists and neuroscientists and often informs computer and animation design. I invoke it here to mark the unease some writers and artists feel when GPT-3 appears to approximate natural language and speech (Caballar 2019).
- 7 For an in-depth discussion of the many debates over the problems with categorization, particularly its history and impact on social identities, see Elam 2022.
 - Many scholars have also challenged different problematic aspects of categorization and classification so central to visual processing systems and versions of Imagenet, a pioneering visual database that categorizes objects, including faces, thereby enabling visual object recognition. Initially created for use in research, some of its commercial and government applications, including surveillance, have come under intense criticism (see Crawford 2021).
- 8 In a keynote lecture Ruha Benjamin (2021) in fact suggests we "embrace the friction"—friction as opposed to the seductive opiate of frictionlessness central to marking technological products but that masks social frictions. She critiques in particular the minimalist design enabling frictionlessness as an aesthetic ideal that intentionally guides users away from the values embedded in a product's infrastructure, from the corporate interests

- animating its design, from extractivism of human labor, and from a tech product contributing to ecocide by its cost to the environment through its making (see Reich et al. 2021, chap. 1).
- 9 Adams quoted in Katz 2020: 6, in a discussion of AI notions of the self: "Practitioners in the 1970s, for instance, offered visions of the self as a symbolic processing machine. . . . In the late 1980s and early 1990s, by contrast, the prevailing 'self' started looking more like a statistical inference engine driven by sensory data. But these classifications mask more fundamental epistemic commitments. Alison Adams has argued that AI practitioners across the board have aspired to a 'view from nowhere'—to build systems that learn, reason, and act in a manner freed from social context. The view from nowhere turned out to be a view from a rather specific, white, and privileged space."
- 10 Personalization is an industry term referring to data-scraping personal information as consumers use their product as part of the business model to better serve client interests and preferences (preferences that of course they then cultivate and curate, and/or sell to third parties if not regulated).
- 11 The "quantified self" is associated with the measuring all aspects of the body and associated with technological self-tracking, lifelogging, quanti-biometrics and auto-analytics. It is often associated with "knowing oneself"—that is, numerically at least—and popularized by wearable fitness and sleep trackers and baby monitoring (see, e.g., Béchard 2021). It showcases a geneticist's efforts to track every single aspect of his body and humans more generally. In the interest of health care, one can only wonder at earlier highly problematic impulses in history to measure humankind also in the name of science, cogently documented in Stephen Jay Gould's field-changing *Mismeasures of Man* ([1981] 1996).
- 12 *Blessings of scale* refers to the observation that for, deep learning, hard problems are easier to solve than easy problems—everything gets better as it gets larger (in contrast to the usual outcome in research, where small things are hard and large things impossible). See Branwen 2022b.
- 13 This contemporary historical moment, in which these modalities are enshrined and embraced in standard technological practice, is eerily similar to an earlier vogue, what Martha Banta (1993: jacket) calls the "efficiency craze" in American culture at the turn of the last century. Banta's fascinating book *Taylored Lives* explores "scientific management: technology spawned it, Frederick Winslow Taylor championed it, Thorstein Veblen dissected it, Henry Ford implemented it. By the turn of the century, practical visionaries prided themselves on having arrived at 'the one best way' both to increase industrial productivity and to regulate human behavior" (jacket).

I am distinguishing between optimization and the many ways in which AI can support the human experience. Though beyond the scope of this article, there are many emerging scientific investigations of the effects of art on the brain: neuro-aesthetics is providing evidence-based research documenting how arts engagement improves brain development and

- cognition, including executive function arts. For a recent example in the area of neuro-aesthetics, see the NeuroArts Blueprint collaboration between Johns Hopkins University and the Aspen Institute: https:// neuroartsblueprint.org. There are also data demonstrating that public arts, mitigating the social alienation inimical to human well-being, is all the more essential to engage in during crises such as war or pandemics.
- 14 These are among many fierce critiques from both within and outside of the tech industry. Note the public controversy over the firing of Timnit Gebru from Google over her research that identified bias in large foundational models such as GPT-3 (Simonite 2021). See also Bender et al. 2021.
- 15 Although Erik Brynjolfson (2022) does not address the issue of art, he also points to the limitations of what he calls the "Turing trap."
- 16 For instance, Branwen and Presser (2019) suggest that "poetry is a natural fit for machine generation because we don't necessarily expect it to make sense or have standard syntax/grammar/vocabulary, and because it is often as much about the sound as the sense. Humans may find even mediocre poetry quite hard to write, but machines are indefatigable and can generate many samples to select from, so the final results can be pretty decent."
- 17 The award-winning documentary Paul Robeson: A Tribute to an Artist (1979) highlights the singer-actor-activist's revisions of the song, originally in the 1921 musical Showboat, over his lifetime from a post-Reconstruction-era melody to a pointed political commentary on racial and economic injustice.
- 18 For an extended discussion of the politics of representing speech, see Elam 1991.
- There is a vast and expanding body of rich scholarship on Black vernaculars and literary representation. For seminal and essential work on this issue, see Gates (1988) 2014.
- 20 See the image at Christie's 2018.
- I place the restoration of *The Night Watchman* in a certain class of AI applications that attempt to recreate, approximate, or better understand the making of an original through the use of AI or other sophisticated technologies. This work is important and much needed, even if it does not necessarily pose challenges to core assumptions and canonized practices in the art world. See, e.g., Stork 2021 and Cann et al. 2021.
- Benjamin articulates the idea of aura as something integral to an original artwork that cannot be reproduced (he was thinking of photography).
- 23 For one of the best accounts of *Birth of a Nation* and how the development of film technology is inextricably tied to the formal encoding of racism in the early twentieth century, see Rogin 1985.
- On the question of humanness and how technologies determine full humans, not quite humans, and nonhumans, see Weheliye 2014. See also Wynter 2003, which critiques that the overrepresentation of Man (as white, Western) as the only imaginable mode of humanness, overwriting other ontologies, epistemologies, and imaginaries (see also McKittrick 2015).

A major influence on this article, Sylvia Wynter's pioneering and prolific work draws on arts, humanities, natural and neuroscience, philosophy, literary theory, and critical race theory. As but one example of the equation of poesy with humanity, consider Thomas Jefferson's (1784–5) infamous argument about the innate inferiority of Black people as the basis for denying them emancipation, that he "never yet could I find that a black had uttered a thought above the level of plain narration; never see even an elementary trait, of painting or sculpture. . . . Among the blacks is misery enough, God knows, but no poetry." Whites were so convinced that a person of African descent was incapable of poesy that when Phillis Wheatley became the first African American to publish a book of poetry, she had to have her master and a gaggle of white officials in Boston testify as proof she was the author.

- 25 Roland Barthe's ([1967] 2001) influential essay "The Death of the Author" critiques the incorporation of biographical background and authorial intent in literary criticism and interpretation.
- 26 For examples of this genre of anxiety, see, e.g.., Crowe 2020, Branwen 2022a, Metz 2020, Elgammal 2019, and Manjoo 2020. See also Pranam 2019 and Offbeat Poet 2019, a discussion of POEMPORTRAITS as "an evolving collective poem generator created by Google Arts and Culture's Ross Goodwin and artist, Es Devlin." The AI is able to implement a creative writing algorithm trained on data consisting of 20 million words from nineteenth-century poetry. One pauses, however, at the goals of art entertainment apps, like the LACMA apps that align your face with a famous painting, which involves data-scraping personal features to add to its database. Is there a privacy disclosure? Is that really an aesthetic experience? And the marketing POEMPORTRAIT opens with an encouragement to donate your word to the making of a collective poem. That language of donation suggests that adding the prompt of a word or line is a way to contribute to some vague collective effort toward an even vaguer social good.
- 27 This refrain appears in many of scholar Ruha Benjamin's public talks, as well as in her books, including *Race after Technology* (Benjamin 2019).
- 28 First coined in a technological context by John von Neumann, *singularity* is meant to describe a point in time when technological advances become inexorable, irreversible, and uncontrollable and cause unforeseeable changes in society. It is often described as a positive possibility in technology; I use it here with a cautionary intent.
- 29 I am drawing on the common definition of *literary signifying* outlined in Gates (1988) 2014.
- 30 In a *Daedelus* issue on AI and society, James Manyika's (2022) afterword similarly reprints experiments with GPT-3. See also the Wordcraft Writers Workshop, a collaboration between Google and professional writers experimenting with co-writing with LaMDA to explore the "rapidly changing relationship between technology and creativity," as described on its landing page https://wordcraft-writers-workshop.appspot.com.

31 There are extant examples of other modes of authorship that do not reflect this more dominant mode of possessive individualism, including many historical African American expressive forms, such as the spirituals, blues, or work songs, that have most often been collective and frequently anonymous works. Indigenous artist-technologist Amelia Winger-Bearskin (2020) also contrasts what I critique as the obsession with genius (almost without exception through history as white/male/cis), especially in the tech world, with what she calls wampum.code ethics.

References

- Baldwin, James. (1979) 1998. "If Black English Isn't a Language, Then Tell Me What Is." In *Baldwin: Collected Essays*, edited by Toni Morrison, 780– 83. New York: Library of America.
- Banta, Martha. 1993. Taylored Lives: Narrative Productions in the Age of Taylor, Veblen, and Ford. Chicago: Univ. of Chicago Press.
- Barthes, Roland. (1967) 2001. "The Death of the Author." *Contributions in Philosophy* 83: 3–8.
- Béchard, Deni Ellis. 2021. "Body Count: How Michael Snyder's Self-Monitoring Project Could Transform Human Health." *Stanford Magazine*, December. https://stanfordmag.org/contents/body-count.
- Bender, E. M., T. Gebru, A. McMillan-Major, and S. Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" FAccT 21. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–23. https://doi.org/10.1145/3442188.3445922.
- Benjamin, Ruha. 2019. Race after Technology: Abolitionist Tools for the New Jim Code. Medford, MA: Polity Press.
- Benjamin, Ruha. 2021. "Which Humans: Reimagining the Default Settings of Technology and Society." Keynote Kieve Lecture at the conference "Antiracist Technologies for a Just Future," Center for Comparative Studies in Race and Ethnicity, Stanford University, Stanford, CA, May 20.
- Benjamin, Walter. (1935) 1969. "The Work of Art in the Age of Its Technological Reproducibility." In *Illuminations: Essays and Reflections*, edited by Hannah Arendt, translated by Harry Zohn. New York: Schocken Books.
- Branwen, Gwern. 2022a. "GPT-3 Creative Fiction." Gwern.net, February 10. https://www.gwern.net/GPT-3.
- Branwen, Gwern. 2022b. "The Scaling Hypothesis." Gwern.net, January 2. https://www.gwern.net/Scaling-hypothesis.
- Branwen, Gwern, and Shawn Presser. 2019. "GPT-2 Neural Network Poetry." Gwern.net, October 29. https://www.gwern.net/GPT-2.
- Brynjolfson, Erik. 2022. "The Turing Trap: The Promise and Peril of Humanlike Artificial Intelligence." *Daedalus* 151, no. 2: 279–94.
- Caballar, Rina Diane. 2019. "What Is the Uncanny Valley?" *IEEE Spectrum*, November 6. https://spectrum.ieee.org/automaton/robotics/humanoids/what-is-the-uncanny-valley.

- Cann, George H., et al. 2021. "Recovery of Underdrawings and Ghost-Paintings via Style Transfer by Deep Convolutional Neural Networks: A Digital Tool for Art Scholars." Paper presented at "Electronic Imaging 2021: Computer Vision and image Analysis of Art," January.
- Chow, Andrew R. 2021. "NFTs Are Shaking Up the Art World—But They Could Change So Much More." *Time*, March 22. https://time.com/5947720/nft-art/.
- Christie's. 2018. "Is Artificial Intelligence Set to Become Art's Next Medium?"

 December 12. https://www.christies.com/features/a-collaboration-between

 -two-artists-one-human-one-a-machine-9332-1.aspx.
- Chun, Wendy Hui Kyong. 2006. *Control and Freedom: Power and Paranoia in the Age of Fiber Optics*. Cambridge, MA: MIT Press.
- Crawford, Kate. 2021. Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence. New Haven, CT: Yale Univ. Press.
- Crowe, Lana. 2020. "Poetech: Shall I Compare Thee to GPT-3? Shakespeare v AI." *Sayre Zine*. https://apostrophezinecom.wordpress.com/2020/10/05/gpt-3-ai-poetry-shakespeare-sonnet-18/.
- Deoras, Shristi. 2020. "GPT-3 Has Weaknesses and Makes Silly Mistakes: Sam Altman, OpenAI," *Analytics India Magazine*. https://analyticsindiamag.com/gpt-3-has-weaknesses-and-makes-silly-mistakes-sam-altman-openai/.
- Drucker, Joanna. 2009. Speclab: Digital Aesthetics and Projects in Speculative Computing. Chicago: Univ. of Chicago Press.
- Elam, Harry J., Jr. 2006. *The Past as Present in the Drama of August Wilson*. Ann Arbor: Univ. of Michigan Press.
- Elam, Michele. 1991. "Dark Dialects: Scientific and Literary Realism in Joel Chandler Harris' Uncle Remus Series." *New Orleans Review* 18: 36–45.
- Elam, Michele. 2022. "Signs Taken for Wonders: AI, Art and the Matter of Race." Issue: *AI and Society*, edited by James Manyika. *Daedalus* 151, no. 2: 198–217.
- Elgammal, Ahmed. 2019. "AI Is Blurring the Definition of Artist." *American Scientist* 107, no. 1: 18–21.
- Gates, Jr., Henry Louis. (1988) 2014. The Signifying Monkey: A Theory of African American Literary Criticism. Oxford: Oxford Univ. Press.
- Gould, Stephen Jay. (1981) 1996. The Mismeasures of Man. New York: Norton.
- Hayles, N. Katherine. 2021. "Narrative and Database: Natural Symbionts." In "Remapping Genre," edited by Wai Chee Dimock and Bruce Robbins. Special issue, *PMLA* 122, no. 5: 1603–8.
- Jefferson, Thomas. 1784–5. *Notes on the State of Virginia*. Volume 1, Chapter 15 "Equality," Document 28. University of Chicago Press. http://press-pubs.uchicago.edu/founders/documents/v1ch15s28.html.
- Katz, Yarden. 2020. Artificial Whiteness: Politics and Ideology in Artificial Intelligence. New York: Columbia Univ. Press.
- Low, Tobin. 2021. "The Ghost in the Machine." Interview with Vauhini Vara. *This American Life*, December 31. https://www.thisamericanlife.org/757/the-ghost-in-the-machine.

- Manjoo, Farhad. 2020. "How Do You Know a Human Wrote This?" New York Times, July 29. https://www.nytimes.com/2020/07/29/opinion/gpt-3-ai -automation.html.
- Manyika, James. 2022. "Afterword: Some Illustrations." Daedalus 151, no. 2: 372–79. https://www.amacad.org/publication/afterword-some-illustrations.
- Marche, Stephen. 2021a. "The Chatbot Problem." New Yorker, July 23. https:// www.newyorker.com/culture/cultural-comment/the-chatbot-problem.
- Marche, Stephen. 2021b. "The Computers Are Getting Better at Writing." New Yorker, April 30. https://www.newyorker.com/culture/cultural-comment /the-computers-are-getting-better-at-writing.
- Mattei, Shanti Escalante-De. 2021. "Artificial Intelligence Restores Mutilated Rembrandt Painting." ArtNews, June 23. https://www.artnews.com/art -news/news/rembrandt-ai-restoration-1234596736/.
- McKittrick, Katherine. 2015. Sylvia Wynter: On Being Human as Praxis. Durham, NC: Duke Univ. Press.
- Metz, Cade. 2020. "Meet GPT-3. It Has Learned to Code (and Blog and Argue)." New York Times, November 24. https://www.nytimes.com/2020 /11/24/science/artificial-intelligence-ai-gpt3.html.
- Mozilla Pulse. n.d. "Amelia Winger-Bearskin." https://www.mozillapulse.org /profile/3119.
- Narang, Sharan, and Aakanksha Chowdhery. 2022. "Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance." Google AI Blog, April 4. https://ai.googleblog.com/2022/04 /pathways-language-model-palm-scaling-to.html.
- Offbeat Poet. 2019. "Putting the ART in ARTificial Intelligence." *Medium*, July 3. https://medium.com/offbeat-poetry/putting-the-art-in-artificial -intelligence-742d6880c34a.
- Paul Robeson: A Tribute to An Artist (Documentary short IMDb 1979). Directed by Narrated Saul J. Turell. Produced by Jessica Berman and Saul J. Turell. Narrated by Sidney Poitier.
- Pranam, Aswin. 2019. "Putting the Art in Artificial Intelligence: A Conversation with Sougwen Chung." Forbes, December 12. https://www.forbes.com /sites/aswinpranam/2019/12/12/putting-the-art-in-artificial-intelligence -a-conversation-with-sougwen-chung/?sh=4c7afa543c5b.
- Ramesh, Aditya. 2021. "DALL-E: Creating Images from Text." OpenAl.com, January 5. https://openai.com/blog/dall-e/.
- Reich, Rob, Mehran Sahami, and Jeremy M. Weinstein. 2021. System Error: Where Big Tech Went Wrong and How We Can Reboot. New York: Harper Collins.
- Rogin, Michael. 1985. "The Sword Became a Flashing Vision': D. W. Griffith's." Representations, no. 9: 150-95.
- Schneider, Tim. 2020. "A.I. Has the Potential to Change the Art Business— Forever. Here's How It Could Revolutionize the Way We Buy, Sell, and See Art." Artnet News, March 31. https://news.artnet.com/market/ai-art -business-intelligence-report-2020-1812288.

- Simonite, Tom. 2021. "What Really Happened When Google Ousted Timnit Gebru." Wired, June 8. https://www.wired.com/story/google-timnit-gebru -ai-what-really-happened/.
- Stork, David. 2021. "Automatic Computation of Meaning in Authored Images Such as Artworks: A Grand Challenge for AI." ACM Journal on Computing in Cultural History 1-10.
- Vara, Vauhini. 2021. "I Didn't Know How to Write about My Sister's Death—So I Had AI Do It for Me." Believer Magazine, August 9. https://believermag .com/ghosts/#content.
- Weheliye, Alexander G. 2014. Habeas Viscus: Racializing Assemblages, Biopolitics, and Black Feminist Theories of the Human. Durham, NC: Duke Univ. Press.
- Winger-Bearskin, Amelia. 2020. "Indigenous Wisdom as a Model for Software Design and Development." Mozilla, October 2. https://foundation .mozilla.org/en/blog/indigenous-wisdom-model-software-design-and -development/.
- Wynter, Sylvia. 2003. "Unsettling the Coloniality of Being/Power/Truth /Freedom: Towards the Human, after Man, Its Overrepresentation—An Argument." CR: The New Centennial Review 3, no. 3: 257–337.