Research Article

Presence in Language Learning Models

Adam Katz
Quinnipiac University

DOI: 10.37514/DBH-J.2023.11.1.02

I would have liked to have seen the late David Bartholomae, in his manner of appropriating high theory for composition studies, bring to bear on Steven Knapp and Walter Benn Michaels' (1982) "Against Theory" his penetrating gaze into the ways interpretative and critical theories tend to presuppose, while suppressing, the specifically pedagogical conditions of entering the academic discourse in which we, as writing faculty, are assumed to be always already conversant. I wonder whether he might have pointed out that if we apply to student writing Knapp and Michaels' axiom that "the meaning of a text is simply identical to the author's intended meaning" (p. 724), we would have to conclude that the student's intention is, first of all, to fulfill an assignment, and, second, to do so in a way that will earn her a satisfactory grade in the class. In that case, working with the explicit statements and arguments in the piece of writing might get us to an "intention" and a "meaning," but could we say either is the student's? The intention evident "in" the text is a simulated or projected one—what the student imagines this teacher will count as a "strong argument" or instance of "critical thinking" (or whatever disciplinary buzzwords the instructor has included in the rubric)—and the instructor will want to attend to that projected intention only insofar as it provides clues to the strategies the student has adopted in attempting to replace the commonplaces she has brought into the academy with some approximation to those current within the academy. It is in that "interlanguage" (Bartholomae, 2005) that we see evidence of learning, which is what we're interested in in the classroom. Knapp and Michaels could probably work with such an intention—after all, they do acknowledge the possibility of a range of speakers from which we'd have to choose, and they do lay down a very minimal notion of intention that is hard to deny: "[w]e know. . . that the speaker intends to speak" (p. 726). Still, I think there's a problem here. If we put it to the student in those terms—something along the lines of "in attempting to mimic the way academic discourse looks to you, you reached for these academic-sounding commonplaces while not noticing these other commonplaces of high school writing, and then stitched up the gaps you noticed by . . ."—the student would not be able to own this as her intention or, except as a possible consequence of instruction that regularly uses these interpretive frames, even be able to see it as a possible intention. And, of course, there are fairly obvious institutional reasons, involving "power relations" (one of those "theory" terms), for the student to refuse to acknowledge acting within an institutional context. So the instructor's reading of the text, even what we could call the "meaning" of the text, might be the "best" one—certainly better, by any measure, than a third party who just tried to figure out what the student was "trying to say" about AI, or whatever—and yet the writer would not recognize it as her own. What kind of intention is it that cannot be recognized as such by the one whose intention it putatively is? It seems to me that this kind of question is the reason we've had something called "theory" in the first place.

I'm starting with these Bartholomaean questions because I'd like to take from *Critical Inquiry* its use of Knapp and Michaels' (1982) essay as a framework for discussing AI. In selecting this essay for the journal's recent forum on AI, the forum's editor, Matt Kirschenbaum (2023), clearly wanted to use the kind of commonsensical understandings of language that theory was created, in large part, to combat to retrieve and regenerate that dispute on what might be auspicious contemporary terms. Still, it might seem an odd and limiting choice since all one can say from Knapp and Michaels' (2023) perspective is that texts produced by ChatGPT and other large language models (LLMs) are not really texts, and we are therefore operating under a delusion (albeit at times, perhaps, a pleasant and harmless one) if we take them to "mean" something. It seems that Knapp and Michael's (1982) evocative example, in their original essay, of the poem produced unintentionally by a wave on a beach was the impetus for taking that essay as the starting point here, but it does constrain the discussion to the narrow question of what kind of "intention," if any, we can ascribe to AI-generated text. We're still thinking in terms of the Turing Test, it seems, where the most interesting question is drawing the line distinguishing humans from computers. But the forum does provide us with examples of more productive approaches, suggestive of what we might do (rather than what we must be careful not to do) with LLM-generated writing, as in the opening paragraph of Ted Underwood's (2023) contribution:

> A graduate student who fell asleep in 1982 and woke up in 2022 might see large language models as a triumph for cultural theory. It is hard to imagine a clearer vindication of a thesis that linguists, critics, and anthropologists spent much of the twentieth century advancing—the thesis that language is not an inert medium used by individuals to express their thoughts but a system that actively determines the contours of the thinkable. (para. 1)

Still, we need not dismiss the question of intention, and what may be Knapp and Michaels' (1982) circular definition opens up some interesting paradoxes and theoretical questions. Their essay, as they acknowledged, is made to fellow literary professionals, with a professional purpose in mind—banishing theory from the profession. But this also means that they see a text as something that "has" meaning and that the relation to texts we are interested in is in interpreting those texts that professional interpreters have found worthy of interpreting. So while they speak of what a text "means," not once do they speak in terms of what a text *says*—a question Bartholomae (2005) recounted posing quite deliberately and insistently to a student in his "Wanderings: Misreadings, Miswritings, Misunderstandings." He insisted on this question because he knew the student would much prefer being asked what the text *meant* since that fits well into the commonplaces students are prepared to serve up ("what Rodriguez is *trying* to say is . . .")—I take Bartholomae to be intervening in the circuit leading from teacher's question to student's answer by presenting the text as an utterance to be responded to rather than an intention to be reiterated "in other words" (that would somehow represent the same intention?). And if "the speaker intends to speak," is not part of that intention to be listened to, to be spoken back to, or to be taken up in some way?

Paul de Man, one of Knapp and Michaels' (1982) theory nemeses, presented a definition of "intention" consistent with this more minimal one of Knapp and Michaels. De Man (1983) distinguished between a commonsensical understanding of intention, wherein

"[i]ntent is seen, by analogy with a physical model, as a transfer of psychic or mental content that exists in the mind of the poet to the mind of the reader, somewhat as one might pour wine from a jar into a glass" (p. 25), and a more phenomenological model, where intent is structural, involving what a type of thing is made for, as a chair for sitting and a text for reading (de Man was far more interested here than I am in the specifically literary text, so I won't follow the particularities of his discussion of aesthetic intentionality). That a text is meant to be read takes us a bit further along than "a speaker intends to speak," but both are consistent with intention determining "the relationship between the components of the resulting object in all its parts" (p. 25) while leaving us free to inquire into those components and their histories and how the institutional form of the object dictates the relationship between the parts. In *Allegories of Reading*, de Man (1982) brought the question into focus in a way that bears forcefully on the pedagogical questions I started with:

> One should not conclude that the subjective feelings of guilt motivate the rhetorical strategies as cause determines effects. It is not more legitimate to say that the ethical interests of the subject determine the invention of figures than to say that the rhetorical potential of language engenders the choice of guilt as theme: no one can decide whether Proust invented metaphors because he felt guilty or whether he had to declare himself guilty in order to find a use for his metaphors. Since the only irreducible "intention" of a text is that of its constitution the second hypothesis is in fact less unlikely than the first. The problem has to be left suspended in its own indecision. But by suggesting that the narrator, for whatever reason, may have a vested interest in the success of his metaphors, one stresses their operational effectiveness and maintains a certain critical vigilance with regard to the promises that are being made as one passes from reading to action by means of a mediating set of metaphors. (pp. 64–65)

De Man's "irreducible intention" matches Knapp and Michaels' (1982) "intention to speak" but is more applicable to problems of writing instruction insofar as constituting a text is not identical to speaking and closer to what we are asking of students. It is likewise undecidable whether the student writer presents himself as "interested" in and "engaged" with a particular topic because he is, in fact, interested and engaged or is deploying tropes whose valuation is, as best the student can determine, attributed to the actual reader of the student's text. (And we can raise the same question about any text, written by anyone, for any audience, which deploys tropes of "interest" so as to constitute a text in the way that, within a particular institutional setting, texts seem to be constituted.)

So I can agree with Knapp and Michaels' (1982) identification of intention with meaning and still, say, insist on one being positioned within a set of institutional relations in such a way that one can produce what counts as a "text" within those relations while still having more or less partial knowledge of how the text has been composed or what a reader might see there. At one end of a continuum, we might have a kind of pure mimicry, the attempt to derive and enact what, to an outsider to a particular discourse, seems to establish the boundaries of that discourse, and at the other end, a deployment of the means of text constitution that marks the historical and institutional consequences of the text and

its circulation familiar to a privileged insider. Both would be equally intentional, but an understanding of the text in terms of the intention of the writer could not be advanced very far by a paraphrase. And, as pointed out, I think, by Bartholomae's "dialogues" between first-year student reading and writing and the reading and writing done by academics so skilled and established that they could afford to forget how they got that way, the sheer intent to engage some discourse in an institutionally privileged way entails considerable unknowing of what we could still call the "meaning" of our texts. (A Wittgensteinian might ask whether we do, in fact, use "intention" and "meaning" synonymously. Why do we need both words? If someone else knows our intent better than we do, is it still *our* intent?)

If we take "intention" in the minimal sense given here, of "speaking to be heard" and "writing to be read," we can develop the notion of intention in a way that will prove revealing regarding the algorithmically mediated machine learning directed at LLMs (and, perhaps, AI more generally). I'll first point out, and will develop soon, the implication that these very simple models of intention are essentially open-ended, coming, we could say, with no necessary expiration date. To say that one speaks to be heard also means that one speaks so that what one says to others might be repeated by others who will in turn repeat (and revise) that and so on. This model is clearly more so the case with writing, which is almost invariably done for readers one doesn't and will never know or interact with, within a broader field of textuality that one might be hoping to modify in ways one could only partially anticipate, A novelist's intention, for example, could be to innovate regarding what can be done with the novelistic form, and such an intention would include a set of assumptions regarding the constitution of novelistic form, its history and variations, forms of institutionalization and canonization, the boundaries separating and connecting it to other forms of textuality, and so on. The intent here might be to provoke in one's readers precisely such a reconstruction, which would lead to a new disciplinary space of inquiry into the novel and which takes us quite a distance from "interpreting" a text so as to "understand" what the author "means"—even if we can still, in a preliminary way, speak in such terms simply by saying that "X is here provoking us into a reconstruction of our received history of the novel." Indeed, once we realize that the initial, and most voracious, readers of texts (what Rhea Myers [2023] called a "paradigmatic audience" [p. 145]) are now algorithms designed to organize, classify and search databases, we can expect writers to increasingly compose works "that [reflect] their ego, or at least [address] them directly (Myers, 2023, p. 145). Myers discussed various ways one might write for the algorithms, some of which would be unintelligible in terms of familiar reading practices (like much innovative writing, for that matter) but highly "meaningful" for the algorithms. For example:

> Below is a text which, to an algorithm, will appear critical of the Digital Humanities, created using negative AFFINN words and the words from Wikipedia's "Digital Humanities" article:
>
> > Ugly despairs racist data lunatic digital computing digital digital humanities digital victim furious horrific research text racism loath computing text humanities betrayed digital text humanities whitewash computing computing cheaters brainwashing digital

> research university university research falsifypo pseudoscience research university worry research . . . (p. 145)

Writing would involve using algorithmically generated text to influence the algorithms that will generate more text (and presumably influence the workings of search engines) in ways that will have a "traditional" end reader such that we could still speak of an intention but one that leads us to a reconstruction of an intervention in contexts rather than to the student reading of a canonical work, which Knapp and Michaels (1982) implicitly took as their model for discerning intention.

So from our minimal definition of "intention," we can get to an application of that definition to forms of textuality very different from those presupposed by professional readers interpreting privileged texts in order to arrive at the correct reading. I would propose calling this understanding of intention "linguistic presence," drawing upon Eric Gans's (2019) *The Origin of Language*, as Gans's use of the concept enables us to more precisely theorize its basis as the constitutive contact between sign-users. Gans used the concept of linguistic presence to solve a very specific problem in unfolding his hypothesis on the origin of language. Gans started with a framework, derived from the literary scholar and cultural theorist René Girard, that places the problem of mimetic rivalry at the origin of human culture, which is to say, the human itself. Girard assumed that what characterizes what we could call "proto-humans" is a high degree of mimetic capacity, which, since we learn how and what to desire from imitating others, leads to conflicts between imitators and their models over possession of mutually desired objects. Girard saw the resolution of the resulting crisis as the "scapegoat mechanism," whereby a single member of the group is singled out by the rest and upon whom their mimetic, resentful "energy" is "discharged." Here Girard located the origin of human sacredness and ritual. Gans approached the mimetic crisis threatening the group differently. He hypothesized that as the group approaches the shared object of desire and begins to see the novel confrontation looming, a sign is emitted by one, then several, and then the entire group in what he called a "gesture of aborted appropriation." More simply put, the movement of the hand to grab the object is converted into a pointing gesture indicating a deferral of appropriation, creating the first sign and referent. Gans then defined human representation as the "deferral of violence," with a different set of implications regarding human culture and history than those that follow from Girard's scapegoating model.

For Gans (2019), the first human sign was "ostensive"—pointing to an object, affirming our shared acknowledgment of its existence as an object that we will refrain from attacking and consuming (and which is therefore available for shared "contemplation"). The problem for Gans, then, was how to get from this originary, ostensive sign to full-blown language or, more precisely, the declarative sentence, which enables us to refer to and discuss things that are not present. So, he constructed a hypothetical sequence of linguistic forms leading from the ostensive to the imperative, then the interrogative, and finally the declarative. I will focus only on one point in this sequence, that taking us from the ostensive to the imperative. Let's pose the problem this way: How could it have been possible for sign-users within an emergent human community to "invent" a new linguistic form like the imperative without, of course, wanting or "intending" to do so since, by definition, they don't know such a linguistic form can exist. (This is, I will suggest, a question we might pose to any form of cultural invention.) Gans imagined this happening through the

"inappropriate" use of the ostensive. That is, someone, perhaps a less experienced sign-user (we could imagine other scenarios), issues an ostensive sign, which is to say, names the object without the object being present. Someone else fetches the object, making the ostensive sign "felicitous." We now have a new linguistic form, one which involves referring to an absent object so that one's listener can retrieve it. In other words, that sequence can now be repeated intentionally, making it an autonomous linguistic form. But why would that listener have responded to the inappropriate ostensive by fetching the object—how would he have "known" to do that? Gans's answer was that the listener is seeking to (*intending* to) maintain linguistic presence, that is, to maintain the scene upon which language is possible because there is an object all parties can point to; and the reason for wanting to maintain linguistic presence is that if the scene collapses, the mimetic violence language was created to defer looms—which also means that what is retrieved as well is the "originary scene" of language's invention/discovery. So, a new cultural form is created "intentionally," but without anyone wanting to create or knowing they were creating that cultural form.

In this case, the concept of linguistic presence provides us with an understanding of language that is intrinsically social and, I will now try to show, well suited not only to describing the language learning undergone by LLMs but to addressing the increasingly urgent problem of embedding these technological developments in an "intentional" human community—a problem a book like Jonathan Roberge and Michael Castelle's (2020) edited volume, *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*, is especially concerned with. Aaron Mendon-Plasek's (2020) contribution, "Mechanized Significance and Machine Learning: Why it Became Thinkable and Preferable to Teach Machines to Judge the World," traces the emergence of pattern recognition as the dominant mode of developing machine learning, precisely because of the contextual significance it presupposes, which furthermore presupposes a community of inquirers within a broader community. Pattern recognition acknowledges the partiality and contingency of knowledge and the need to incorporate feedback, which can be scaled up or down as necessary. This reliance on feedback makes pattern recognition convergent with the cybernetic displacement of "representation" with a more participatory mode of soliciting knowledge through interaction:

> What could people do using pattern recognition that they couldn't do before? What made pattern recognition's problem framing not merely rhetorically compelling but intellectually preferable for some communities? Which communities celebrated and were empowered by these capacities? Pattern-learning machines offered a way of imperfectly knowing the world via its provisional and piecemeal traces. Mid-century pattern recognition shared with mid-century cybernetics what Andrew Pickering called a "black box ontology," in which the world is filled with black boxes "that [do] something, that one does something to, and that does something back," and for which its inner workings are opaque to us. Pattern recognition systems, like the cybernetic systems Pickering discusses, attempted to "go on in a constructive and creative fashion in a world of exceedingly complex systems" that might never reasonably be understood or derived from first principles. (p. 40)

In this case, machine learning is inextricable from human learning, suggesting the possibility of an inquiry-based, interactive approach to LLMs in which those models would "ask" us to identify patterns in a preliminary way, and then resume pattern recognition activity modified by ongoing human intervention, in the course of which the "machine" would, increasingly often, identify patterns that its human "collaborators" would not have but now can.

Pattern recognition is essentially the same as what I am referring to as "linguistic presence," where we (or a computer program) can look at two separate marks, let's say A and *A*, and, within a particular community with shared tacit understandings, identify both as the Latin letter "A." In other words, the most minimal understanding of "intention," and now "linguistic presence," is being able to say "this is the same"—what I see is what you see (or hear, or touch, etc.). Being able to "confirm," or "affirm," or "authenticate," or "acknowledge," that "this is the same" reiterates the basic human gesture of pointing to something, what the primatologist Michael Tomasello (2010) called "joint attention" and showed, as simple as it is, not to be part of the communicative repertoire of our closest relatives, the great apes. Linguistic presence is impossible for apes because they don't have scenes, and joint attention can only take place on a scene: we need to be configured so as to be oriented toward some object of joint attention. Once we have the concept of linguistic presence to say "this is the same" on a scene, we can extend and scale up our understanding of what can be a "scene," as needed. After all, what is a laboratory if not a scene upon which our senses are distanced and enhanced through various devices so that some very thoroughly trained "we" can detect together some conversion or transformation and can "register" and "measure" something that would have been unavailable otherwise. And what is our entire technoscientific civilization if not a product of such scenes and a means of producing more of them, installing them in other "scenes," and making them more penetrating, sensitive and precise, so that we can find more objects, infinitesimal as well as enormous, beyond the capacity of our unaided senses to grasp, about which we can say, in many different ways, "this is the same"?

Linguistic presence, then, can already be stretched across a multitude of scenes at varying scales and "inter-scenic" articulations, but it is the very scenic nature of "presence" that makes automating it problematic: It hardly needs to be said that we can identify particular "objects" as the same ("certify" them, so to speak) without them being, in any "objective" (extra-scenic?) sense, the same. What counts as a difference or distinction depends upon what we're looking for, and what we are looking for is conditioned by what our own presence on a particular scene, nestled within a range of other scenes, primes us to look for. So, while the "same/other" binary might be constitutive of the digital, as Alexander Galloway (2022) argued in a recent essay, "The Golden Age of Analog," Galloway also reminded us that the analog, based on "likeness," is still very much with us. And couldn't we see "likeness" as a relaxation of the rigor of "sameness"—anything that is the same as something else (even itself) in some, maybe many, respects is going to be different in other respects, and insofar as we acknowledge the differences, we can "stretch" or "transition" "same" into "like" and then realize that, as Paul North (2021) argued at great length in his *Bizarre-Privileged Items in the Universe: The Logic of Likeness*, everything is like everything else in some respect. And it turns out that representing the world in terms of degrees and modes of likeness, as those degrees and modes might be ascertained from within as wide a range of different scenes as possible, is a far better way of increasing the

intelligence of our interactions with computation than representing the world as an exhaustive sequence of same/other circuits with which it is nevertheless continuous.

If you show anyone two objects and ask whether they are "like" each other, you can always get an answer. You will get one kind of answer if you specify in which respects the objects are to be considered, and how precisely, and another kind of answer if you just let the person choose whatever criteria they wish. Here we have the difference between supervised and unsupervised learning, as you proceed to automate these designations (Wasielewski, 2023). If someone marks 100 pairs of objects as alike or unlike, a programmer can write an algorithm to have a computer "look at" another 1,000 pairs and determine whether that person would have been likely to consider each pair to be alike or unlike. Unsupervised learning will approximate supervised learning, or, at least, start to look like supervised training, insofar as the machine, in reproducing the logic of the choices made by the subject, will be identifying "features" of the different objects that the subject "seemed" to be using as criteria for her decisions. But the unsupervised learning will certainly be more diverse and idiosyncratic. If we make as our goal increasingly intelligent interactions with computation, the simplest way of accomplishing that is to provide "intelligence" of the ways human interaction is informing the "trajectory" of the machine learning. If we "stay in touch" with the program, we can intervene periodically simply by saying, "yes, I would mark as 'like' what the program has," or "no, I wouldn't," and review the results accordingly. Now, this is the kind of work done by Amazon's "Mechanical Turks" that has become a persistent object of critique by AI-skeptics (in particular, AI-hype opponents), and as a form of employment involving assessing decisions made by others on topics in which one is uninterested or finds repulsive, it would be highly tedious, unpleasant and even traumatic. But if it's your own work, things might be very different. (The synthesis of supervised and unsupervised learning is creating categories of likenesses that emerge through unsupervised learning and using them to supervise learning going forward.)

I am trying, here, to provide a sketch of a model for engaging critically with AI within the broader logic of data collecting (including scraping) and analysis and the broader sensing and measuring infrastructure that, as Benjamin Bratton (2016) has informed us, is engirding the earth. There is a broader, one might say, "political" implication of the model I've presented so far, which looks, intentionally, like a model for data exchange. As is well known, some of the most contentious issues that have arisen regarding not only LLMs but AI more generally include the automation of surveillance and security systems as well as the collection and use of data in unaccountable, even unlawful, and abusive ways, which Shoshana Zuboff (2019) described as "surveillance capitalism" and what Katherine Bode and Lauren Goodlad (2023) saw in "Data Worlds: An Introduction" as a new kind of "primitive accumulation." I'll also note that one of the proposals that seems to come up periodically (since the early 2000s) to deal with unwarranted data collection is to have companies that do so pay users of a platform for their data. This proposal, which is usually dismissed quickly as "utopian" (perhaps rightly, under contemporary conditions), already presents a model of data exchange. Perhaps an example of where the benefits of such data exchange are easier to see can point us toward further discussion. We can all see why it would be beneficial for the medical institutions with which we will all interact throughout our lives to be able to gather data from us as patients and use that data, which may require for its value more or less contextual information about those who have supplied it, for

future studies into disease identification, prevention and treatment. Here, both sides of the exchange are clear: they get the data they need to do their work while we get improved healthcare, assuming we trust the medical institutions to serve their founding purpose (which, of course, I acknowledge we may not be able to). With the LLMs that have become the main topic of AI discussion more recently, what users get in return is less clear (although many are certainly finding ChatGPT and other programs useful); and, moreover, how such data exchange is to be formalized so that there is some commensurability between what one gives, as an individual and as a member of a community (or several overlapping communities), and what one receives, will need to be addressed. That's what the model I'm working out here aims at facilitating.

We start to implement a data exchange model once we start writing for the algorithm as our primary "audience" because then we are more likely to receive a commensurate "packet" of data, in large part because we (progressively, provisionally) know what we are looking for. In other words, we are teaching the program while learning alongside it. One of the questions raised regarding the design and ownership of LLMs is access to algorithms themselves, including the weights given to various tokens, and this is certainly a matter of data exchange. But eliciting results through purposefully designed and iterated prompts will surface, if not the actual "mechanics" of the program (assuming non-engineers would know what to do with that), then the elements providing for increasingly targeted estimates of what the program takes to be alike or the same. In this way, one is implementing a kind of data exchange program unilaterally and breaking with a humanist model of sign exchange between individual humans, which may not exist anymore and perhaps never did. The scene upon which one initiates a gesture others might follow in saying "this is the same" is simply prolonged, perhaps indefinitely, but one acts in the meantime by producing collateral gestures soliciting responses from others in such a way as to increase the likelihood of the "closure" of that scene. In other words, your writing becomes a program for your own continued pedagogical practice, as we can say that "learning" is simply re-positioning yourself on a scene and reconfiguring the scene in such a way that those upon the scene could say "this is the same" regarding some matter that they would have been unable to identify previously. A simple example of this is looking through a microscope or telescope and needing someone else to tell you what you are "looking at," thereby enabling you to participate on the scene of observation and inquiry. Thus, setting the terms of an ongoing data exchange, or at least making one "bid" after another, does not require any knowledge whatsoever of programming language; rather, it requires a cultivation of a new set of aesthetic sensibilities, a new range of ways of seeing how things might be the same as or like each other. The condition, though, would certainly be an acceptance of expanded, mediated or "stacked" scenes, which is to say an acknowledgment that "meaning" and "intention" will not reach closure in the individual interaction between interlocutors or reader and writer—which was really the lesson of "theory" all along.

The distinction between machine learning based on pattern recognition, or, in Mendon-Plasek's (2020) words, "mechanical schemes for imitating human judgment" (p. 42), even when you don't yet know what you're looking for, and computational text generation that attempts to model symbol manipulation on the cognitive, analyzable operations of the mind finds its equivalent in the difference between a pedagogy based on successive approximation to a target ("average") discourse and a pedagogy based on discrete skills that can presumably be isolated and identified and assessed separately. It's

the difference between, say, identifying the distinctive vocabulary of a particular text against the background of more familiar vocabularies and then collaboratively using that vocabulary to name the moves undertaken as readers of that text, on the one hand, and identifying a list of "attributes" of "critical thinking" ("supporting claims with evidence," "showing cause-and-effect relations," "identifying the assumptions underlying claims," etc.) which leaves the student with no recourse other than to guess what the teacher takes to be "valid evidence," a "causal relationship" or the assumption underlying a particular claim, on the other hand. Nothing about asserting that one thing causes another, or one claim "supports" the proof of another claim, enables one to generalize about causality or evidence in general, while learning how to use another's language helps one to construct a practice of using another's language on other occasions. Do that enough, and under conditions where the language doesn't quite "fit," and it becomes "your" language, and it will be possible to show students when that happens (you will be constituting scenes upon which likeness and sameness can be acknowledged). Similarly, engaging with what a programmed and trained language model predicts you will say next makes explicit one out of a range of things you might possibly say next, and you can learn how to "tilt" the model towards suggestions that follow more closely the predictions you consider more worth pursuing. That the program is not really "thinking" or "communicating" with you is only a matter of concern if you're working with a model of language use bound to identifying the intention behind the meaning of the words presented to you—in that case, you will feel cheated or deceived. But if you see yourself as establishing linguistic presence with an unknown range of others (among whom you might count some future version of yourself) at a distance mediated by the totality of linguistic exchanges, then you need only be concerned with operating on the algorithm in such a way as to generate "pedagogical platforms" out of its various outputs and in that way creating more favorable conditions for data exchange. If the program is black-boxed, so are humans, and we can, for machines and humans alike, simply look for markers of learning as the staging of a mimetic approximation to an "average" of some database, aimed at increasingly lowered thresholds of same/other distinctions. All we can do is examine what we do and study the ways doing one kind of thing (what, as a disciplinary community, we provisionally deem to be a "kind of thing") can be articulated with other kinds of doing so as to perform in ways expected within the spaces that distinguish between the "kinds of things" in question. And pursuing those expected ways of doing things beyond the expected ways of pursuing them is what produces the unexpected, around which we could need to reconfigure "likeness" and "sameness."

The equivalent of object recognition programs for writing would require the classification of writings into types, which the writer engaging with AI would identify along same/likeness lines as part of doing research into the histories of discourses in their institutional situations. So, one could prompt an LLM to produce a situation comedy written by Poe, creating an imaginary object of comparison against which you would measure a particular product as the same or like. As you proceed, the assumptions and expectations regarding an increasingly wide range of discourse would be explored, and research projects into, say, something like topical humor in Poe's poetry would be constructed and pursued. A community of inquirers into, say, Poe or comedy might enter this process with vague notions of perhaps underappreciated comic elements in Poe's writing or that his writing is devoid of humor. These assumptions, in turn, implicate other assumptions regarding humor (how to identify and explain it) and Poe, as a writer in a

historical situation and processed through histories of interpretation, and these assumptions might vary widely across the disciplinary space in ways that no one could have articulated or even considered beforehand. Each sample provided by the model can be engaged with the simple question "Does this strike you as a Poe-like script for a sitcom?" along with some preliminary reasons for your answer, and then revisions to the prompt following new characterizations of Poe's work, or of comedy, or what might be comic in Poe's work, etc., would progressively surface these assumptions in the process of providing new answers. Insofar as contemporary large language models don't allow for such inquiries, enabling them to do so can be at the top of our list of demands to present to their makers.

Lauren Goodlad (2023), in her introduction to the new journal *Critical AI*, "Humanities in the Loop," argued for parity with and collaboration between discourses within the humanities and the emergence of AI (with the humanities already transformed by the digital humanities). Near the end of her discussion, Goodlad gestured towards the fracturing of the humanities over the past half-century, raising the question of how equipped they are as an interlocutor with what some are beginning to call not "artificial" but "amplified" intelligence:

> I conclude with a parting word to another group among *Critical AI*'s potential readers. If you, dear reader, regard yourself as a humanist of some stripe—perhaps a literary critic, historian, political theorist, philosopher, or digital humanist—the invocation of "humanities" discourse may strike you as strangely belated. After all, would not the "humanist" readers of a new interdisciplinary journal recognize themselves and their most cogent ideas as, by now, *post*human in every conceivable way—as fragmented, reassembled, and distributed as many digital processes? As commodified and datafied as any late-capitalist artifact? As stripped of any pretense to biological or cultural privilege as the barest of bare life?

Where, then, does the necessary "critical" standpoint situate itself? After the thoroughgoing deconstruction and dismantling of the humanist subject along the philosophical and political lines (by "deconstructionists, feminists, postcolonial theorists and critical race scholars") presupposed by *Critical Inquiry*'s "Again Theory" forum, where do the humanities stand? Goodlad suggested a kind of answer, and perhaps a new mission for the humanities, precisely in the anthropomorphic illusions generated by AI's imitation of human intellectual practices:

> With respect to "AI," critical perspectives perceive how anthropomorphic analogies misrepresent the functionalities of data-driven machine systems when they conflate predictive analytics with human decision-making and equate massive datasets with human knowledge, social experience, and cultural commitments. The point of rejecting such flawed assumptions is as much to capture robust understandings of machine intelligence as it is to complicate mechanistic simplifications of biological life. (Reductive and Controversial Meanings of "Intelligence" section)

Goodlad here constructed a couple of boundaries that it would, presumably, be the vocation of the humanities to examine and defend: between "predictive analytics" and "human decision-making," "massive datasets" and "human knowledge, social experience, and cultural commitments." These will be moving boundaries, especially since the purpose of improving predictive analytics is to improve human decision-making and massive datasets combined with increasingly comprehensive search programs to register and transform human knowledge, social experience and cultural commitments. Further inquiry into central concepts in literary studies, and aesthetics more generally, would reveal ways in which those vocabularies are always already permeated with technology and, in particular, writing. (For just one example of how much of this work is already taking place, see Daniel Shore's (2018) *Cyberformalism: Histories of Linguistic Forms in the Digital Archives*.) Most importantly, to critique "anthropomorphic analogies," one must distinguish what is "human" from what is only "like" the human (and upon particular, purposefully treated and curated scenes). Doing that in turn means that the humanities must return to the most basic of all of its questions: What is the human? I suggested an answer above in my discussion of Eric Gans's (2010) originary hypothesis and will return to the point with a more canonical representative of the 20th-century humanities, Kenneth Burke, who also proposed a hypothesis of the origin of language quite consistent with ("like") that of Gans. In his "A Dramatistic View of the Origins of Language and Postscripts on the Negative," Burke (1966) hypothesized that the first word must have been a "negative," more precisely the kind of negative we find in the "admonitory": *don't* . . . . Gans constructed a more precise account by presenting a more carefully constructed scene of origin than the dramatistic Burke, but I would admonish all of us to heed their identification of the origin of language and the human in a shared danger that we pose to ourselves and that we must defer. Burke posed a kind of originary indebtedness binding us to each other:

> Yet the mention of private property brings up another point. We have already indicated, and shall later consider more fully, how moral negatives can become positives through universalization. For if everybody were in debt to everybody, to this extent nobody would owe anybody. At least, the indebtedness would cancel out. So far as sheer mathematics is concerned. But we must consider a twist whereby the genius of the moral negative, as thus made positive, can add a new kind of negativity, in the very midst of its positivizing. For if everybody has something that he would keep for himself to the exclusion of everybody else, to this extent everybody is guilty with regard to everybody, so that the accumulation of such positive possessions adds up to universal indebtedness. (p, 434)

No AI can be present on such a scene, can share such a debt. We "anthropomorphized" ourselves before, and as a precondition of, becoming vulnerable to further "anthropomorphic analogies." I'm putting forward, necessarily contentiously, a "universal" proposal for the humanities, but one that could be infinitely particularized as the forms of "debt" and "repayment" are worked out, alluding but irreducible to the forms of financialization constituting our institutions presently. If the ever permeable and shifting boundaries constituting the human are to be maintained as a condition of criticality

regarding anthropomorphic analogies, then it will be through the "scheduling" of data exchanges as, at least, "installments" that new forms of reciprocity might be generated.

References

Bartholomae, D. (2005). *Writing on the margins: Essays on composition and teaching*. Palgrave MacMillian.

Bode, K., & Goodlad, L. M. E. (2023). Data worlds: An introduction. *Critical AI,1*(1). https://doi.org/10.1215/2834703X-10734026

Bratton, B. (2016). *The stack: On software and sovereignty*. MIT Press.

Burke, K. (1966). *Language as symbolic action: Essays on life, literature and method.* University of California Press.

de Man, P. (1982). *Allegories of reading: Figural language in Rousseau, Nietzsche, Rilke and Proust*. Yale University Press.

de Man, P. (1983). *Blindness and insight: Essays in the rhetoric of contemporary criticism*. University of Minnesota Press.

Galloway, A. R. (2022). The golden age of analog, *Critical Inquiry, 48*(2), 211–232. https://www.journals.uchicago.edu/doi/full/10.1086/717324

Gans, E. (2019). *The origin of language* (New ed.). Spuyten Duyvil.

Goodlad, L M. E. (2023). Humanities in the loop. *Critical AI, 1*(1–2). https://doi.org/10.1215/2834703X-10734016

Kirschenbaum, M. (2023, June 6). Again theory: A forum on language, meaning, and intent in the time of stochastic parrots. *In the Moment.* https://tinyurl.com/bddrpczr

Knapp, S., & Michaels, W. B., (1982). Against theory. *Critical Theory, 8*(4), 723–742.

Knapp, S., & Michaels, W. B. (2023, June 30). Here is a wave poem that I wrote . . . I hope you like it! *In the Moment*. https://critinq.wordpress.com/2023/06/30/here-is-a-wave-poem-that-i-wrote-i-hope-you-like-it/

Mendon-Plasek, A. (2020). Mechanized significance and machine learning: Why it became thinkable and preferable to teach machines to judge the world. In J. Roberge & M. Castelle (Eds.), *The critical life of machine learning: An incursion into critical AI studies* (pp. 31–78). Palgrave MacMillan.

Myers, R. (2023). *Proof of work: Blockchain provocations 2011–2021*. Urbanomic.

North, P. (2021). *Bizarre-privileged items in the universe: The logic of likeness*. Zone Books.

Roberge, J., & Castelle, M. (2020). *The critical life of machine learning: An incursion into critical AI studies*. Palgrave MacMillan.

Shore, D. (2018). *Cyberformalism: histories of linguistic forms in the digital archive*. Johns Hopkins University Press.

Tomasello, M. (2010). *Origins of human communication*. MIT Press.

Underwood, T. (2023, June 9). The empirical triumph of theory. *In the Moment*. https://critinq.wordpress.com/2023/06/29/the-empirical-triumph-of-theory/

Wasielewski, A. (2023). *Computational formalism: Art history and machine learning*. MIT Press.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.