

Solomon Katz Distinguished Lectures in the Humanities
at the University of Washington

FEB 10
2026

Emily M. Bender

**Resisting Dehumanization in the Age of “AI”:
The View from the Humanities**

**SIMPSON
CENTER for the
HUMANITIES** ✦
COLLEGE OF ARTS & SCIENCES
UNIVERSITY of WASHINGTON

Resisting Dehumanization in the Age of “AI”: The View from the Humanities

Emily M. Bender
University of Washington

Katz Distinguished Lecture in the Humanities
10 February 2026

@emilymbender.bsky.social



“AI” research, development and sales involves dehumanization on many levels

- Computational metaphor
- Digital physiognomy
- “Ground lies”
- Irrelationality
- Ghost work
- Reinforcing the white racial frame

What can humanities scholars do about this?

- With all others: speak up about environmental impacts and exploitative labor practices; be skeptical consumers
- Resist the logic of domains (see Ribes et al 2019)
- Resist hierarchies of knowledge (see Gebru 2021)
- Resist calls to use “AI” in our teaching and scholarship
- Make vivid what it is that we do — and its relationship to text & image
- Make vivid what it is to inhabit various identities, and to experience relationality

Roadmap

- Definitions & researcher stance
- Dehumanization: working definition
- Dehumanization in “AI” research, development & sales
- What humanities scholars can do

Definitions

- “AI” doesn’t refer to a coherent set of technologies, but has always been a marketing term
- Term was coined in 1955 by John McCarthy as part of a grant proposal for a summer project
- You’re inundated with it now, not because of tech advances but because of enormous investment by VC and other interests
- In this talk, I am using it to refer to:
 - The collection of products sold as “AI”
 - The ideology that automating “intelligence” is possible and valuable

Researcher stance/Who am I?

- PhD training in syntax and sociolinguistics
- Long experience with multilingual grammar engineering: building grammars in software, across (mostly spoken) languages
- Since 2016: methodologies for supporting consideration of societal impacts of language technology—in NLP research, development, and education.
- Broader conversation about identifying and mitigating harms done in the name of “AI”

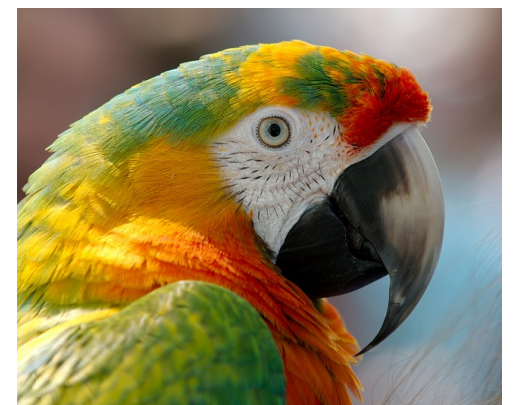
Climbing towards NLU: On Meaning, Form and Understanding in the Age of Data (Bender & Koller 2020)

- Written in reaction to widespread claims that language models “understand” language
- But language is a system of signs (pairing of form & meaning; de Saussure)
- Language models (GPT-3 et al) are trained on the task of string prediction: their only input is form
- Comparisons to child language acquisition are misleading: child learn language in socially rich, socially situated interactions
- Octopus thought experiment: posit an intelligent learner, given access only to form; all that is learned is patterns in form



On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 (Bender, Gebru et al 2021)

- Observed a trend towards ever larger language models, and asked:
- What are the possible risks associated with this technology and what paths are available for mitigating those risks?
- Environmental costs & environmental racism
- Financial costs & impact on research participation
- Datasets filled with hegemonic viewpoints & worse; no/minimal documentation and accountability
- Synthetic text generating machines can reproduce systems of oppression
- Synthetic text generating machines can mislead humans



But how do I know that
you're not just a
stochastic parrot?

Dehumanization: Definitions

- “**Dehumanization happens when people are depicted, regarded, or treated as not human or less human.** [...] I start with such a thin notion since not much agreement exists beyond it in the scholarship on dehumanization, not even with respect to the above examples. Most scholars will count them as dehumanizing, while others will not.” (Kronfeldner 2021:xvii)
- “If racialization is understood not as a biological or cultural descriptor but as a conglomerate of sociopolitical relations that discipline humanity into full humans, not-quite-humans, and nonhumans, then blackness designates a changing system of **unequal power structures that apportion and delimit which humans can lay claim to full human status and which humans cannot.**” (Weheliye 2014:3)

Dehumanization: Working definition

1. Cognitive state of failing to perceive another human as fully human
2. Acts that express that cognitive state or otherwise entail the assertion that another human is not fully human
3. Experience of being subjected to acts that express lack of perception of one's humanity and/or deny human experience or human rights

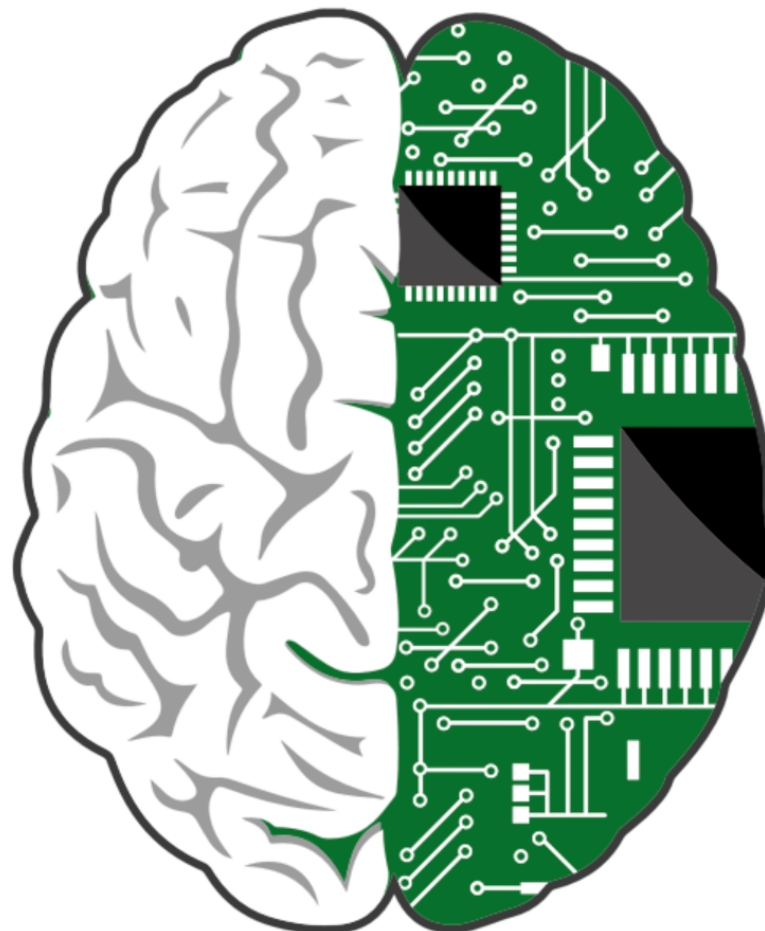
Fully human

- Entitled to all rights recognized as human rights
- Equally in possession of an internal life and point of view
- Welcomed and known as one's full self

Dehumanization in the research,
development and sales of “AI”

The Computational Metaphor (Baria & Cross 2021)

- Scientific metaphor used and debated in neuroscience: THE BRAIN IS A COMPUTER
- PR metaphor used by technologists: THE COMPUTER IS A BRAIN



The Computational Metaphor (Baria & Cross 2021)

- “afford[s] the human mind less complexity than is owed, and the computer more wisdom than is due.” (p.2)
- “the Computational Metaphor rests on other well-ingrained ideologies in which a hierarchy of human value is **tied to a particular notion of intelligence** such that the quality of **being emotional is considered inferior to being rational.**” (p.6)
- “This notion of intelligence extends to the justified subjugation of beings considered less rational to those considered (or propagandized as) more rational, whether animals to humans, women to men, or one race of humans to another. According to this logic, **in its fake-ness as a human intelligence, AI paradoxically succeeds in being a more trustworthy form of intelligence**, by being the epitome of rational thought.” (p.6)

On anthropomorphism in science (Dijkstra 1985)

- “A more serious byproduct of the tendency to talk about machines in anthropomorphic terms is **the companion phenomenon of talking about people in mechanistic terminology**. The critical reading of articles about computer-assisted learning [...] leaves you no option: in the eyes of their authors, the educational process is simply reduced to a caricature, something like the building up of conditional reflexes. For those educationists, Pavlov’s dog adequately **captures the essence of Mankind** —while I can assure you, from intimate observations, that **it only captures a minute fraction of what is involved in being a dog—.**”



Appropriating experiences of disabled people to assert humanity of “AI”

- Agüera y Arcas (blog, 12/2021) asserts that LLMs are like Deafblind people
- Under the heading “modality chauvinism” calls on the writings of Daniel Kish and Helen Keller to argue that no one sensory system is required for humans to develop concepts
- But he can’t show that large language models are like people, with internal lives, relationships, and full personhood
- The analogy ends up dehumanizing Blind and Deafblind people, by saying they are like something that is patently not human, specifically because of their disability.

<https://bit.ly/EMB-blog1>

Still finding this argument in 2026

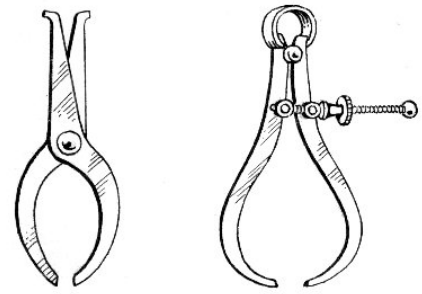
- From Chen et al, comment in Nature, 2 February 2026

They don't have bodies. Without embodiment, critics argue, there can be no general intelligence. This reflects an anthropocentric bias that seems to be wielded only against AI. People would ascribe intelligence to a disembodied alien communicating by radio, or to a brain sustained in a vat. An entity that responds accurately to any question, but never moves or acts physically, would be regarded as profoundly intelligent. Physicist Stephen Hawking interacted with the world almost entirely through text and synthesized speech, yet his physical limitations in no way diminished his intelligence. Motor capabilities are separable from general intelligence.

Important new book: Williams 2025 *Disabling Intelligences*

Digital Physiognomy

(see Agüera y Arcas, Mitchell & Todorov 2017)



- Classification of people into identity categories or personality characteristics based on computer processing of photos, voice signals, or other
 - Claims of predicting: criminality, sexual orientation, employability, political leaning, psychopathy, etc (see Stark & Hutson 2022)
 - Gender, race, etc classification similarly problematic
- Flattens human identities and emotional experiences into intrinsic, externally observable categories of classification
- Pseudoscience of physiognomy made apparently “objective” through the application of computers

“Ground lies”

- Data sets used in training “AI” systems are mythologized as representative, due to size or lack of curation (Paullada et al 2021, Raji et al 2021, Scheuerman et al 2021)
- Decisions at every point: where to collect from, how to collect, how to filter, what labels to apply, who should apply them, how to verify labels
- If we don’t actively work to curate the datasets we want, we *will* be collecting datasets representative of dehumanizing ideologies like white supremacy
- “Data sets so specifically built in and for white spaces represent the constructed reality, not the natural one. **To have accuracy calculated in the absence of my lived experience** not only offends me, but also **puts me in real danger.**” (Raji 2020)

Humans are not just social but thoroughly relational

- “The self thus never just *is* but rather emerges continuously and jointly relying on behavior and action and **on doing and being together with others.**” (Kyselo 2014:8)
- “**Humans are inherently historical, social, cultural, gendered, politicized, and contextualized organisms.** Accordingly, their knowing and understanding of the world around them necessarily takes place through their respective lenses.” (Birhane, 2021:5)
- “The alternative to relativism is partial, locatable, critical knowledges sustaining **the possibility of webs of connections** called solidarity in politics and shared conversations in epistemology.” (Haraway, 1998:584)
- “**Situated knowledges** are about **communities**, not about isolated individuals. The only way to find a larger vision is to be somewhere in particular.” (Haraway, 1998:590)

AI “knowing” is irrelational

- “Data science and data practices reincarnate rationalism in many forms, including [...] the manner in which the dominant view is taken as the “God’s eye view” (Birhane 2021:3)
- Machines aren’t designed to be in relationship, to jointly make meaning, or to apply “*mētis*” (Scott 1998, Alkhatib 2021)
- *Sealed knowledges*: Chatbots as a search interface *de-situate* knowledges (cf. Haraway 1988) and, Lindemann (2025) argues, *seal* them, foreclosing user's chances to engage with information landscapes and promoting a “view from nowhere”.

Irrelationality:

Devaluing humanity while leaving no space for it

- “But there is also an underlying presupposition almost always at play that suggests, tacitly and otherwise, that the **dehumanized and anonymous decision-making** done by computers in a way that mimics—but replaces—that of human actors is somehow **more just or fair.**” (Roberts 2021:52)
- “Damage manifests most profoundly not only when errors get made, but when people are compelled to endure those errors. [...] absurdity follows when algorithmic systems deny the people they mistreat the status to lodge complaints, let alone the power to repair, resist, or escape the world that these systems create. In other words, **when algorithmic systems try to insist that they live in their utopias.**” (Alkhatib 2021:3)

Ghost work:

Humans as hidden software components

- Human effort is everywhere in so-called “AI” systems: data labeling, system design, evaluation, and backstop for when the task is too difficult for the machine
- Tech firms hide the labor and humanity of microworkers in systems designed to produce the illusion of AI
- Crowdwork platforms hide humanity of microworkers from requestors by representing workers only through their worker IDs and selling them as interchangeable
- ... while sometimes appropriating people’s professional prestige: OpenAI boasts that ChatGPT Health was made in consultation with 260 physicians ... who annotated 600,000 model outputs for them.

(Gray and Suri 2019, Roberts 2021)

Learn More: Data Worker's Inquiry

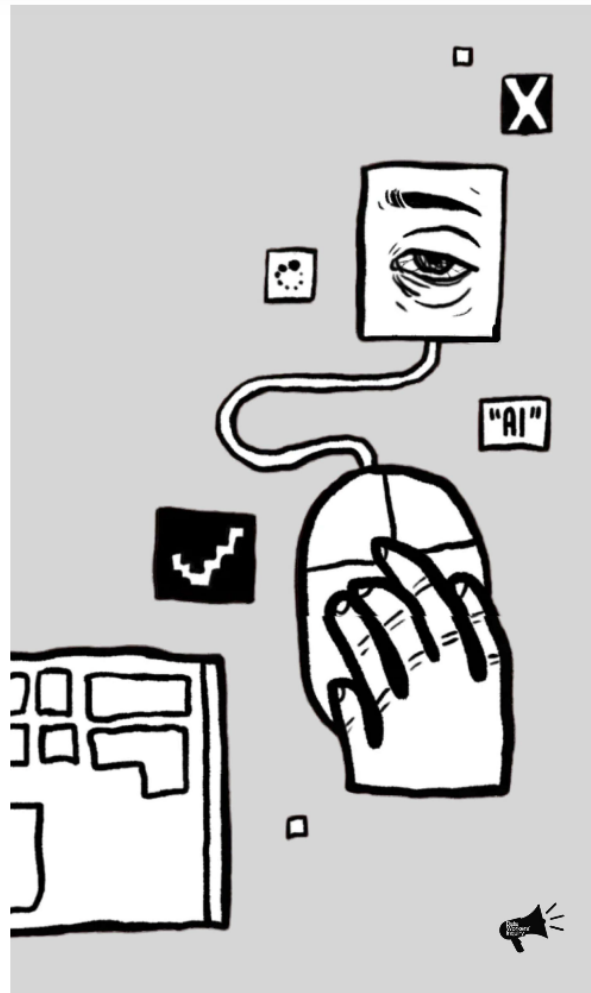
data-workers.org



The Labor Behind AI Intimacy

by Michael Geoffrey Asia

When users think they are talking to a machine, human emotion becomes data.



"If I Had Another Job, I Would Not Accept Data Annotation Tasks"

by Roukaya al-Hammada

This report examines the unique challenges faced by Syrian refugees working as data annotators in Lebanon.



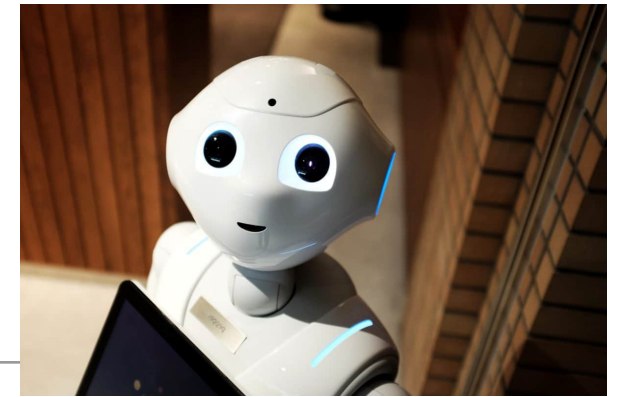
The Impact of Gift Card Payments on MTurk Workers

by Alexis Chávez

Amazon Mechanical Turk pays data workers with gift cards instead of money. This report underlines how this payment method fosters workers' dependency, especially in countries conditioned by financial turbulence.

Reinforcing the white racial frame

(Cave & Dihal 2020)

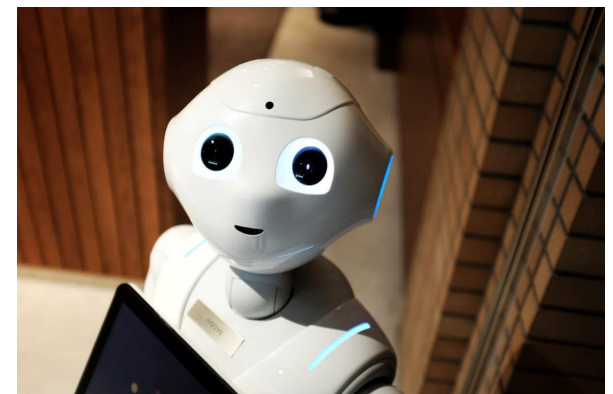


- Within Anglo Western culture at least, AI is racialized as white
 - Robots are frequently literally made with white (color) exteriors
 - More humanoid robots are designed to be perceived as racially white
 - Even voice assistants & text-based chatbots mostly “talk white”
 - Weizenbaum’s ELIZA used white language features (Marino 2014)
 - Siri released in 2011, African American voices for it only in 2021

Reinforcing the white racial frame

(Cave & Dihal 2020)

- Cave & Dihal's hypothesized causes for this:
 - Disproportionately white workforce in AI
 - The traits associated with AI (intelligence, professionalism, power) are those that the white racial frame ascribes to white people
 - “The Whiteness of the machines allows the White utopian imagination to fully exclude people of colour.” (p.698)



What can humanities scholars do about this?

Resist alongside others, in calling out

- Environmental devastation of data centers & especially “AI”
- Data theft & labor exploitation
- Privacy concerns
 - Follow news coverage
 - Raise these issues in discussions (e.g. “AI” at the university)



Key questions to ask, to be a skeptical consumer (of tech, of tech news)

- How is the task defined? What's the input / what's the output?
- Is there any reason to believe that the input provides sufficient information to produce accurate output?
- Where did the training data come from and how was the system validated?
- How were/are the data workers treated?
- Who benefits from assuming the output is accurate?
- Can this technology be used for surveillance, harassment, or otherwise denying people their rights?



Emily M. Bender @emilymbender · Jul 3



Not it **effing** can't. This headline is breathtakingly irresponsible.

h/t [@hypervisible](#)



[bloomberg.com](#)

Algorithm Claims to Predict Crime in US Cities Before It Happens

A new computer algorithm can now forecast crime in a big city near you — apparently.

85

799

3,332





Emily M. Bender

@emilymbender



And now we are squarely in the ML techno-solutionism danger zone: It's established that it would be beneficial to have something that can do X with only Y input, but not that it's actually possible to do X with only Y input.

11:56 AM · Apr 6, 2022 · Twitter Web App

28 Retweets **9** Quote Tweets **163** Likes



Emily M. Bender @emilymbender · Apr 6



Replying to @emilymbender

On the other hand, you can always train an ML system that takes y_s (elements of Y) as input and gives x_s (elements of X) as output and thus LOOKS LIKE it's doing X with only Y input.



1



17



140



Resist the “logic of domains” (Ribes et al 2019)

- AI proponents like to talk about “domain experts” and “domain expertise”
- This is sometimes even respectful, on the surface
- But it flows from the “logic of domains”, whereby CS expertise is domain-independent, general, “objective”
- But expertise in how to automate things is just one kind of expertise, like any other
- It doesn’t sit in some external space; it doesn’t provide a “view from nowhere” (Haraway 1988)

Don't hold space for "AI" that might work elsewhere

- "AI" peddlers are incentivized to dismiss the objections of "domain experts", saying "you're just trying to preserve your job"
 - ... while seeming less vested in any given conversation, since their systems are supposedly general
- If protein folding models, based on carefully constructed datasets have proven useful to biologists and go under the same name ("AI") as the synthetic text extruding machines that are despoiling our info ecosystem, it is harder to criticize the latter with one's full chest.
- Inverting that: Anyone who wants to call what they are doing "AI" should be accountable for making clear why their product (or research) isn't in fact slop.

Resist hierarchies of knowledge (see Gebru 2021)

- Which fields are valued at our institutions?
- What makes the rigor of the fields legible to people across the university?
- How and why does state and federal funding track private interests?
- Challenging hierarchies of knowledge is especially important in deflating AI hype
- The current moment is a culmination of devaluing ways of knowing outside STEM. It's time to past time push back.

Resist calls to use this in our teaching & research

We're going to prepare for this AI future that everyone is talking about by committing to funding fundamental research across disciplines, but especially the humanities and social sciences. Of course, we're concerned about the ethical and equitable development and use of the technology, and that's why we need scholars who are innovating at the edges of our understanding of how humans experience life, how power works in society, and how we can reshape our social and economic systems towards justice, equity, and sustainability. And we recommit to our mission of training students to be critical thinkers across disciplines, who can consider sources of information and locate them within their context, who can evaluate toolkits for the tasks they are taking on and decide which tools fit which task, and who can see through the glib marketing that power cloaks itself in.



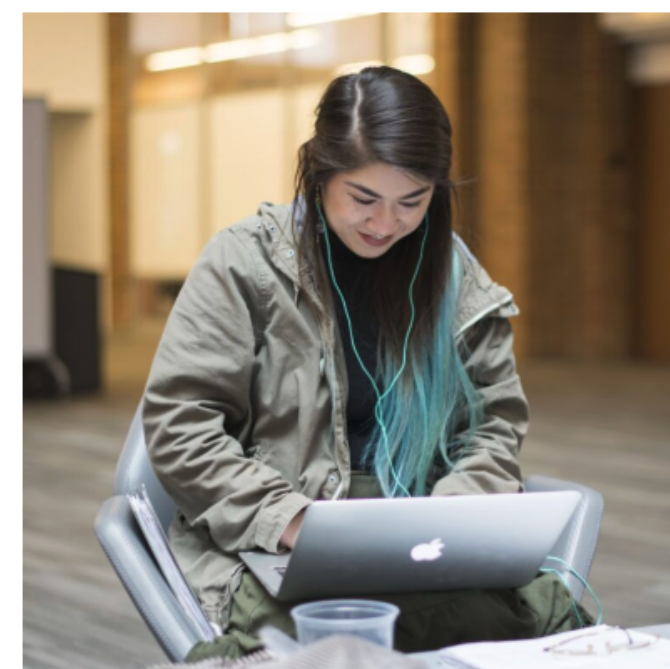
(Bender & Hanna 2025:96)

Resist calls to use this in our teaching and research

AI@UW announces SEED-AI 2026 grants, with awards of up to \$50,000 available

The University of Washington AI@UW Initiative is now accepting proposals for the Supporting Educational Excellence and Discovery with AI Grants (SEED-AI 2026). Faculty and teaching staff can apply for grants ranging from \$1,000 to \$50,000 to fund projects that explore how artificial intelligence can enhance teaching and learning. Applications are due February 1, 2026, with awards announced on or around March 1, 2026.

Funded by the Charles and Lisa Simonyi Launch Fund for Artificial Intelligence, SEED-AI supports innovative projects that apply AI thoughtfully to improve education, inspire future developments, and benefit the UW community. Projects with broad impact, such as those serving large numbers of students or creating resources for faculty, will be prioritized.



Resist calls to use this in our teaching and research

October 23, 2025

Purple: UW's Custom AI Tool Is Taking Shape



Purple, UW's custom AI tool, is moving from concept to reality. And it's already being tested across all three campuses.

Resist calls to use this in our teaching and research

- Drimmer & Nygren 2025: “We believe that the intellectual, ethical, and institutional downsides to using this technology are so substantial that normalizing its integration into pedagogy poses risks that far outweigh whatever benefits one might associate with it. In fact, we would argue that thus far the only benefits to using AI in art historical research have been to demonstrate how poorly equipped it is to conduct research in the historical humanities.”
 - See also Nygren & Drimmer 2023
- Dusseau 2024: “If you are tired of the drumbeat of inevitability that insists English faculty adopt AI into our teaching practices, I am here to tell you that you are allowed to object. Using an understanding of human writing as a means to allow for-profit technology companies to dismantle the imaginative practice of human writing is abhorrent and unethical.”

Resist calls to use this in our teaching and research

- Sano-Franchini et al 2024: “[W]e demonstrate that our call for GenAI refusal is not uninformed, moral outcry as a result of technological “aversion,” fear of change, or other so-called “doom and gloom” views of technological advancement as it is often framed, but rather extends on our disciplinary knowledge in rhetoric, writing, and composition studies, digital rhetorics and literacies, computers and writing, and technical communication.”
- Logan 2024: “If, as a field, we are to realize the conference theme and the call to “strengthen our own practices leading to human flourishing,” we might design learning experiences that help people to map the ecologies of GenAI and its relations as part of a broader Luddite praxis dedicated to collective action against inequities and towards more just futures.”

Resist “anti-progress” or FOMO frame

- Each time we refuse “AI”, we make it easier for others (and ourselves) to refuse again.
- We can also make positive statements about what we value: authenticity, accountability, human connection, and credit where credit is due (e.g. to artists for their work).
- If someone wants to be dismissive of those values, that is a choice they can make.
- By taking the initiative to frame the debate and making clear what is at stake, we can show that choice for the dangerous, anti-social one that it is.

Make vivid what it is that we do

And its relationship to text and image

- The “AI” peddlers would have you believe that what matters is just the text
- And in the humanities, a lot of our work can be anchored in texts
 - That we read
 - That we write
- So it is up to us to articulate what it is that we do and what it means
- ... and how we relate to other people (co-present or otherwise) while doing it

The meaning is not in the text



- What does this sentence *mean*?
- What does the speaker *mean* by uttering this sentence?

先生によると男の子よりも女の子がポケモンがすきだ。 [jpn]

先生 によると 男の子 よりも 女の子 が ポケモン が すき だ。

Sensei ni yoru to otokonoko yorimo onnanoko ga pokemon ga suki da.

teacher (.) boy (.) girl (.) Pokemon (.) like (.)

teacher ACCORDING.TO boy THAN girl NOM Pokemon NOM like COP.PRES



The meaning is not in the text

- With linguistic (grammatical, lexical) knowledge, speakers can get from a text to a ‘standing’ or ‘conventional’ meaning (Grice 1968, Quine 1960), but that’s only the first step.
- Standing meaning + commonsense + coherence relations gives *public commitments* (Hamblin 1970, Lascarides & Asher 2009, Asher & Lascarides 2013)
- Public commitments + further reasoning gives *perlocutionary consequences*
 - A: I wonder whether I should take my umbrella. Is it raining?
 - B: Yes.
 - A: Oh, so you do think I should take my umbrella.
 - B: I didn’t say that.

(Bender & Lascarides 2019:13)

Conversation as a joint activity: Clark 1996 (p37-38)

Participants	A joint activity is carried out by two or more participants.
Activity roles	The participants in a joint activity assume public roles that help determine their division of labor.
Public goals	The participants in a joint activity try to establish and achieve joint public goals.
Private goals	The participants in a joint activity may try individually to achieve private goals.
Hierarchies	A joint activity ordinarily emerges as a hierarchy of joint actions or joint activities.
Procedures	The participants in a joint activity may exploit both conventional and nonconventional procedures.
Boundaries	A successful joint activity has an entry and exit jointly engineered by the participants.
Dynamics	Joint activities may be simultaneous or intermittent, and may expand, contract, or divide in their personnel.

Communication as intersubjective awareness (Baldwin 1995, p.132)

Technically speaking, joint attention simply means the simultaneous engagement of two or more individuals in mental focus on one and the same external thing. Put this way, joint attention is likely a ubiquitous occurrence for all organisms that boast a complex central nervous system. For instance, two bushbabies, alerted by a predator's call, are caught in an instant of joint attention prior to pursuing their separate avenues of escape. Or to take a human case, perhaps you and I once unwittingly happened to watch "Dr. Strangelove" on the same night in the same time zone, thereby satisfying the criteria for joint attention. Clearly, this notion of simultaneous engagement fails to capture something central to our experience—the aspect of intersubjective awareness that accompanies joint attention, the recognition that mental focus on some external thing is shared. And of course, it is just this aspect of the joint attention experience—intersubjective awareness—that makes simultaneous engagement with some third party of such social value to us. It is because we are aware of simultaneous engagement that we can use it as a springboard for communicative exchange.

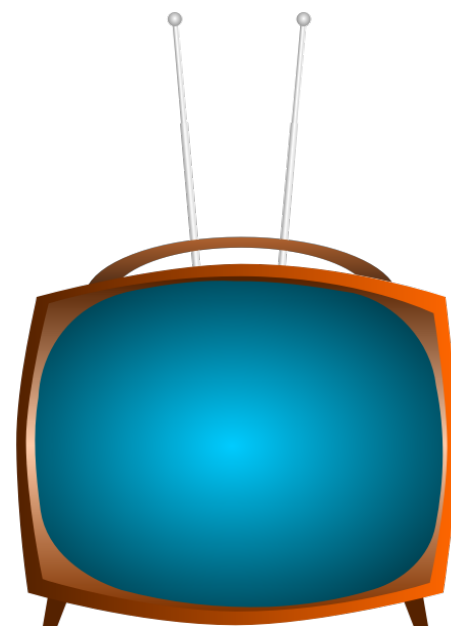
Meaning making at a distance: in time & space



- Face-to-face, small group communication is the most well-studied (and probably the most basic)
 - but we also communicate asynchronously and distantly
 - and apply the same skills in doing so
- Theory of mind developmental milestones linked to reading comprehension (Atkinson et al 2017, Dore et al 2018)
- Ricœur 1973 (hermeneutics): “Not that we can conceive of a text without an author; the tie between the speaker and the discourse is not abolished, but distended and complicated.” (p.95)
- In interpreting texts, we lack the ability to confirm & repair understandings (Dingemanse et al 2015), but we still project a model of mind

Making meaning in human-human interaction: Summary

- Communication is a joint activity
 - in which we use language (among other signals)
 - to convey and understand communicative intents
- We do this even when not co-present with our interlocutors



Photograph by [Rama](#), Wikimedia Commons, Cc-by-sa-2.0-fr

The Great Chatbot Debate



<https://bit.ly/ChatbotDebate>

Make vivid what it is that we do

And its relationship to text and image

- The AI peddlers would have you believe that what matters is just the text
- And in the humanities, a lot of our work can be anchored in texts
 - That we read
 - That we write
- So it is up to us to articulate what it is that we do and what it means
- ... and how we relate to other people (co-present or otherwise) while doing it

Make vivid what it is to live in different identities, and to experience relationality

- Visual arts

- Plays

- Poetry

- Memoir

- Fiction

- Music

- ...

“AI” is ultimately
“technologies of isolation”
-Chris Gilliard

The humanities are about
experiencing “being human
together”
-Gabriel Solis

“AI” research, development and sales involves dehumanization on many levels

- Computational metaphor
- Digital physiognomy
- “Ground lies”
- Irrelationality
- Ghost work
- Reinforcing the white racial frame

Humanities scholars have many ways we can resist

- With all others: speak up about environmental impacts and exploitative practices; be skeptical consumers
- Resist the logic of domains (see Ribes et al 2019)
- Resist hierarchies of knowledge (see Gebru 2021)
- Resist calls to use this in our teaching and scholarship
- Make vivid what it is that we do — and its relationship to text & image
- Make vivid what it is to inhabit various identities, and to experience relationality

References

- Agüera y Arcas, B. (2021). Do large language models understand us? Blog post on Medium.com, <https://medium.com/@blaisea/do-large-language-models-understand-us-6f881d6d8e75>.
- Agüera y Arcas, B., Mitchell, M., and Todorov, A. (2017). Physiognomy’s new clothes. Blog post on Medium.com, <https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a>.
- Alkhatib, A. (2021). To live in their utopia: Why algorithmic systems create absurd outcomes. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–9.
- Asher, N. and Lascarides, A. (2013). Strategic conversation. *Semantics and Pragmatics*, 6(2):2:1–:62.
- Atkinson, L., Slade, L., Powell, D., and Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology*, 164:225–238.
- Baldwin, D. A. (1995). Understanding the link between joint attention and language. In Moore, C. and Dunham, P. J., editors, *Joint Attention: Its Origins and Role in Development*, pages 131–158. Psychology Press.
- Baria, A. T. and Cross, K. (2021). The brain is a computer is a brain: Neuroscience’s internal debate and the social significance of the computational metaphor. <https://arxiv.org/abs/2107.14042>.
- Bender, E. M. (2019). The #benderrule: On naming the languages we study and why it matters. *The Gradient*. <https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters/>.
- Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S., and et al (2021). On the dangers of stochastic parrots: Can language models be too big?  In *Proceedings of FAccT 2021*.
- Bender, E. M. and Hanna, A. (2025). *The AI Con: How to Fight Big Tech’s Hype and Create the Future We Want*. HarperCollins, New York.
- Bender, E. M. and Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.
- Bender, E. M. and Lascarides, A. (2019). *Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics*. Morgan & Claypool.
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2):100205.
- Cave, S. and Dihal, K. (2020). The whiteness of AI. *Philosophy & Technology*, 33(4):685–703.
- Chen, E. K., Belkin, M., Bergen, L., and Danks, D. (2026). Does ai already have human-level intelligence? the evidence is clear. *Nature*, 650:36–40.
- Clark, H. H. (1996). *Using Language*. Cambridge University Press, Cambridge.
- Dijkstra, E. W. (1985, September 25). On anthropomorphism in science. Philosophers’ Lunch, Austin, TX, transcript available at <https://www.cs.utexas.edu/users/EWD/ewd09xx/EWD936.PDF>.
- Dingemans, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladdottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., et al. (2015). Universal principles in the repair of communication problems. *PloS one*, 10(9):e0136100.
- Dore, R. A., Amend, S. J., Golinkoff, R. M., and Hirsh-Pasek, K. (2018). Theory of mind: a hidden factor in reading comprehension? *Educational Psychology Review*, 30(3):1067–1089.
- Drimmer, S. and Nygren, C. J. (2025). Reflections on AI in the classroom. *ICMANEWS: The Newsletter of the International Center of Medieval Art*, Spring 2025, No. 1:25–35.
- Dusseau, M. (2024). Burn it down: A license for AI resistance. *Inside Higher Ed*.
- Gebru, T. (2021). The hierarchy of knowledge in machine learning and related fields and its consequences. Talk presented at the Spelman College Center of Excellence for Minority Women in STEM, April 14, 2021. Recording available at <https://www.youtube.com/watch?v=0L3DowBM9uc>.
- Grice, H. P. (1968). Utterer’s meaning, sentence-meaning, and word-meaning. *Foundations of Language*, 4(3):225–242.
- Hamblin, C. (1970). *Fallacies*. Methuen.
- Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3):575–599.

- Kronfeldner, M., editor (2021). *The Routledge Handbook of Dehumanization*. Routledge, New York.
- Kyselo, M. (2014). The body social: An enactive approach to the self. *Frontiers in Psychology*, 5:1–16.
- Lascarides, A. and Asher, N. (2009). Agreement, disputes and commitment in dialogue. *Journal of Semantics*, 26(2):109–158.
- Lindemann, N. F. (2025). Chatbots, search engines, and the sealing of knowledges. *AI & Society*, 40:5063–5076.
- Logan, C. (2024). Learning about and against generative AI through mapping generative AI’s ecologies and developing a Luddite praxis. In *Proceedings of the 18th International Conference of the Learning Sciences-ICLS 2024*, pages 362–369. International Society of the Learning Sciences.
- Marino, M. C. (2014). The racial formation of chatbots. *CLCWeb: Comparative Literature and Culture*, 16(5):13.
- Nygren, C. and Drimmer, S. (2023). Art history and ai: Ten axioms. *International Journal for Digital Art History*, 9:5.02–5.13.
- Paullada, A., Raji, I. D., Bender, E. M., Denton, E., and Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns*, 2.
- Quine, W. V. (1960). *Word and Object*. MIT Press, Cambridge MA.
- Raji, D. (2020). How our data encodes systematic racism. *MIT Technology Review*.
- Raji, I. D., Bender, E. M., Paullada, A., Denton, E., and Hanna, A. (2021). Ai and the everything in the whole wide world benchmark. In *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks*.
- Ribes, D., Hoffman, A. S., Slota, S. C., and Bowker, G. C. (2019). The logic of domains. *Social Studies of Science*, 49(3):281–309. PMID: 31122173.
- Ricœur, P. (1973). The model of the text: Meaningful action considered as a text. *New Literary History*, 5(1):91–117.
- Roberts, S. T. (2021). Your AI is a human. In Mullaney, T. S., Peters, B., Hicks, M., and Philip, K., editors, *Your Computer is on Fire*, pages 51–70. MIT Press.
- Sano-Franchini, J., McIntyre, M., and Fernandes, M. (2024). Refusing GenAI in writing studies: A quickstart guide.
- Scheuerman, M. K., Hanna, A., and Denton, E. (2021). Do datasets have politics? Disciplinary values in computer vision dataset development. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2).
- Scott, J. C. (1998). *Seeing Like A State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, New Haven, CT.
- Stark, L. and Hutson, J. (2022). Physiognomic artificial intelligence. *Fordham Intellectual Property, Media and Entertainment Law Journal*, 32(4).
- Weheliye, A. G. (2014). *Habeas Viscus: Racializing Assemblages, Biopolitics, and Black Feminist Theories of the Human*. Duke University Press, Durham, NC.
- Williams, R. M. (2025). *Disabling Intelligences: Legacies of Eugenics and How We are Wrong about AI*. Springer Nature.