JOURNAL OF INTERACTIVE MEDIA IN EDUCATION

Unsafe AI for Education: A Conversation on Stochastic Parrots and Other Learning Metaphors ⚠

COLLECTION: METAPHORS OF AI IN HIGHER EDUCATION: DISCOURSES, HISTORIES AND PRACTICES

ARTICLE

EMILY M. BENDER
EAMON COSTELLO
KYUNGMEE LEE
ROBERT FARROW
GISELLE FERREIRA

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

This interview article discusses the impact on popular and educational discourses of the metaphor for a Large Language Model of a "stochastic parrot". The metaphor comes from the title of an influential paper on the harms of large language models from 2021. Here we present a perspective on the short but influential history of the metaphor through an interview with one of its creators, Professor Emily M. Bender. Using the broad lens of metaphor as a way to shape and frame discourse, the editors interviewed Professor Bender and asked her a series of questions to spark discussion around AI in Education. A variety of topics were covered, including: on how metaphors and anthropomorphisation when carelessly used can elide harms and obviate responsibilities; the role of BigTech, data theft and metaphors of colonisation; Whether AI is unsafe for education and if so to which learners; Techno-solutionism; Positionality and AI "voice"; and whether EdTech is a key driver of AI bullshit and enshittification. This article aims to give readers an accessible insight into how a particular metaphor may be enacted in discourse and to contribute to wider critical debates about how GenAI operates in the context of datafication and educational harms.

CORRESPONDING AUTHOR:
Eamon Costello
Dublin City University, IE
eamon.costello@dcu.ie

KEYWORDS:
metaphors; bullshit; enshittification; GenAI; AIED; Edtech; LLM; bias; AI discourse

TO CITE THIS ARTICLE:
Bender, EM, Costello, E, Lee, K, Farrow, R and Ferreira, G. 2025. Unsafe AI for Education: A Conversation on Stochastic Parrots and Other Learning Metaphors ⚠. *Journal of Interactive Media in Education*, 2025(1): 10, pp. 1–16. DOI: https://doi.org/10.5334/jime.1079

> *"Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot."*
>
> (Bender et al. 2021: 616–617)

INTRODUCTION

The stochastic parrot has become shorthand for a range of problems associated with the accuracy of (Large) Language Models (LLM). Often, LLM outputs mask the socio-political, material realities of their production, including harmful environment and labour practices. The term was introduced in a landmark paper by Bender et al. (2021) that has at the time of writing over 8,000 citations attesting to its impact on research and scholarship. The paper had immediate material impacts also. Google's alarm at the paper's critique of its technology is reported to have led to it firing two employees who were paper authors, Timnit Gebru and Margaret Mitchell, and who had also raised concerns about diversity and censorship within the company (The Guardian 2021; BBC News 2021).

Since its publication the metaphor of the parrot has found its way into many popular discourses (Jandrić 2024). It has been namechecked by Bill Gates; tweeted by Sam Altman; featured in headlines from the Wall Street Journal to the Thailand News; and has had companies and products named after it. In 2023, it edged out ChatGPT and LLM to be crowned AI-related Word of the Year by the American Dialect Society (American Dialect Society 2024) appearing as both a noun and a verb.

Accordingly, the editors conducted an interview with Professor Emily M. Bender (EMB), to ask her about the influence of the stochastic parrot metaphor which she co-created. All guest editors of this special collection devised and agreed the questions/themes, shared them with Professor Bender in advance and Eamon Costello (EC), Robert Farrow (RF) and Kyungmee Lee (KL) were part of the synchronous discussion conducted over Zoom. The interview was conducted on 20th of October 2024, was subsequently transcribed and edited before the final text was agreed by all authors.

In the interview we asked Professor Bender about her perspective on how the metaphor of the parrot has evolved, what role metaphors can play in popular discourse and how those discourses can impact key debates around AI in higher education. In the liminal space before we pressed record we chatted briefly about hope via the metaphor of whether the glass is half full or half empty. Whether this piece itself is hopeful or not however will depend partly on your position and perspective as a reader, both on the glass you hold and how you look into it. We hope you enjoy reading.

1. THE LEGACY OF THE STOCHASTIC PARROT

EC: I'm going to ask you about whether you think the parrot has been a force for good in the discourse or not, whether it has been misunderstood and misused? So you have two metaphors from which to choose which could represent you. Are you:

A. A proud, pirate queen whose charge has left her shoulder, and is out there doing its work in the world? Spreading the bad news, squawking inconvenient truths?

Or

B. A concerned, conscientious ornithologist, whose Amazonian escapee is an invasive species that you're trying to get back in its cage?

EMB: Oh, that's an interesting question! I think I have to go for the pirate queen, though. So the purpose of the metaphor, when we initially coined it, was to make vivid the way in which large language models – they weren't even called large language models yet right? – to make vivid the way that they are doing fake language. Language can be thought of as a system for communication, as a system of signs, where you have a pairing of form and meaning. If you

DOI: 10.5334/jime.1079

run large language models in the mode where they are text synthesis machines or synthetic text extruding machines, then they are doing only the form side of language and the point of the metaphor was to make that vivid. I've disliked some things about what's happened since publication. First of all, we use the noun "parrot". So, it sounds like we're referring to the animal, and we used the emoji in the paper title. So we definitely are drawing on the image of a parrot. But the verb "to parrot" is actually a better source for the metaphor. The verb to parrot, meaning to repeat back without understanding sort of sets aside what's going on with actual parrots that undoubtedly have inner lives of some kind. So, that was one problem. The other thing is that it did get picked up as a sort of caricature of the argument by people who strongly identify with so-called artificial intelligence, who are deeply invested in artificial intelligence, and they took the phrase *stochastic parrot* as an insult against this thing that they had identified with. The next step from there is that people turn around and use it as an insult against other people.

EC: You say that it's a verb, but what about the suggestion that it's not actually parroting back? It could be creating novel outputs: if I type something into Google, it gives me back the verbatim text. Isn't the parrot creating something "novel" all the time?

EMB: Well, actually, if you type something into Google these days, you get that AI overview, which is a stochastic parrot thing, right? It's unbidden (speaking of AI intrusion). But to your point about not getting verbatim quotes from the training data, that's what stochastic means. Stochastic is randomised, but according to the probability distribution, so it's parroting things back, but not verbatim.

RF: One of the things I think of with the parrot metaphor is that it can imply that AI can do things that it can't – because of the metaphor. So, for instance, to use the parrot idea: it almost confers the idea that this is as intelligent as a parrot, or something like that, that it has something like the agency of a parrot, something like the awareness of the world of a parrot and that kind of thing. Obviously, it doesn't have those things, but there's a sort of mental sleight of hand. To give another example is the hallucination metaphor, as people will say: "Oh it's hallucinating, that's all" when a GPT has given you a load of rubbish. To me, hallucination kind of implies conscious experience, which can be true or false, right? Because hallucination is a false, conscious experience. But obviously, that's not the case with AI. It's not having any experiences right? But you can be led to believe this thing is hallucinating, that it's capable of hallucination. So do you think, with the parrot metaphor, or with others, that there's been something unhelpful or misleading?

EMB: Definitely agree with you there on hallucination. I think that it's a problematic term, because it suggests that the large language model is perceiving something, right? Hallucinations are false perceptions, as I understand it, but there's no perception at all. That's one layer of the problem. The other is that hallucinations can be symptoms of serious mental illnesses. And so using the word sort of in this joking way is, I think, problematic in an ableist fashion.

And the third point is that it makes it sound like mostly it's right, but *sometimes* it gets things wrong, right? It's trying, right? There's some anthropomorphization there that says sometimes it makes a mistake when in fact, it's always just working with "what's a likely next word". And if that corresponds to something that is true, or is a useful thing to say, then that's by chance. When we wrote the paper, we quite liked the metaphor, and so it ended up in the title. But the particular part of the paper that motivates the phrase – stochastic parrots – we thought we were kind of on thin ice there, because at the time in late 2020, for the most part, people were using language models not to generate text, but as components of larger systems. This is the sort of original use case for language models. If you think about automatic transcription, you have a system that goes from the sound sequence to possible word forms. And one component of that is a language model that re-ranks those sequences of word forms as – okay, this looks more like the training corpus – this is what's more likely to have been said – *because* it looks more like the training corpus.

So, the ultimate source of the information is not the language model itself, but whoever was creating that audio signal; same thing with machine translation. You have a sequence of words in one language, a translation model that gives you a range of possibilities in the target language, and the language model sort of smooths that out. That was the main use

of language models at the point that we were writing the paper. But OpenAI had started with GPT-2 and GPT-3 as ways of generating text.

We said, "Okay, well, if people start doing that, that's going to be problematic because other people are going to encounter that language and think that it's coming from a person." People are going to think that it's coming from someone who has accountability for what it's saying; but it's actually just haphazardly stitching words together from this training data.

So that's where that was then, and, in the years since, it has become a much bigger deal right? Now people are not only creating synthetic text and putting it in front of other people to see, but in fact, we are encouraged to go to the synthetic text extruding machines as sources of information. Many people are doing that. So at that point, I guess you could say, yes: stochastic parrot is problematic – especially because people don't really know what stochastic means, right?

EC: That's something that's quite interesting, because it's quite nice to have an image of a bird, and people talking about animals.

RF: It's also the idea of a natural creature being the metaphor, right? Nature! And here's this AI thing, which, okay, it's debatable the extent to which human creations are part of nature right? That's a philosophical debate, but the metaphor immediately implies this kind of *organic thing*.

EC: And also you said Emily, that there are certain people who identify with AI who got really upset with the parrot, you know, and this kind of idea that we think of ourselves as humans, but we don't really interrogate what that is.

EMB: Yeah, so I think that when people are using "stochastic parrot" as an insult, they're using it to mean it doesn't have any thoughts of its own. I think the people who are using it as an insult want to believe that these large language models, as "AI" systems, are really thinking and have ideas. And some people want to think they're sentient. So they take "stochastic parrots", the phrase, to mean it has no inner life, which is true. That's what we mean by the phrase. And it's not an insult to say that about a large language model, because it's just a description, but it is a very insulting, dehumanising thing to say about people.

RF: I'm a philosopher by background, and earlier in the summer, I was back in my old university, where I did my PhD and I was talking to one of the professors about AI. I was saying to him that it only took me a while to understand this whole thing about one token after another, the predictive model. You know, here's the next letter. Here's the next word. He responded, "well, that's kind of how I do it right when I'm thinking: One word after the next." Now... he was joking right? But is it so different when we don't really understand our own cognition? Do we know that we're not doing something similar?

EMB: The difference that I'm going to insist on as a linguist is that when we experience language, we are always experiencing both the form and the meaning. And in fact, it's really hard to see that they're different because of that experience. When you're experiencing linguistic input, in a language that you are competent in, you are always immediately seeing both form and hypotheses about the meaning. What was the person trying to convey with that?

What's going on with large language models is form only. Their training data is literally sequences of letters. You put enough text in, and the systems come out with this ersatz fluency that makes it seem like there's a whole lot more in there. So, whatever mapping between text and image is truly in the model, we are going to believe that it's more because we do so much sense-making when we see the linguistic part of it. I have a whole paper called "Resisting Dehumanization in the age of AI" (Bender 2024), which started partially because Sam Altman tweeted "I am a stochastic parrot and so r u " right? And he's basically saying, "yeah, well, humans are no better than that." And, to me, that is minimizing what it is to be human with the goal of lifting up these systems so he can make bigger claims about them.

RF: Yeah, and make more money ultimately.

EC: Can I ask something about that? I mean, I don't know why this is relevant, but one thing you've done in your paper, that is very nice, and I don't know whether it was a factor in its success, but there's a playful element to it – certainly in the title – to what you're doing. It has an almost raucous feel to it, in some sense, even though it's a really rigorous, scholarly piece.

Even down to the emoji, you know, and I think that feels like a kind of a resistance – using metaphors and deploying things like this and the parrot. It's a kind of a reaction to this utter kind of seriousness.

EMB: I think there are probably two sources to that. One is, I belong to a scholarly tradition in linguistics that has a history of playfulness and not taking itself too seriously. My training is in formal linguistics, syntax, grammar. I did some work in sociolinguistics, too, and especially the syntacticians, the people that I studied with, Ivan Sag and others, were very quick to see humour and acknowledge that we're not curing cancer here, right? It's a privilege to be able to spend time thinking about the structure of language. And if we take ourselves too seriously, that's going to be to the detriment of the scholarship. So I think that's part of it. There are lots of playful paper titles in the field that I come from!

EC: Epistemic humility in what you're doing there that's lacking in the discourse. In a lot of the discourse, the people that frame the discourse, I think.

EMB: Yeah. So that's one source and then the other thing is that I have been running a podcast – and we started live streams in 2022 – called Mystery AI Hype Theatre 3000. And one relevant idea due to my co-host, Alex Hanna (also co-author of my forthcoming book) is "ridicule as praxis", and the idea is that sometimes one of the most effective things you can do in combating problematic things in the world is to show the humor in it. We are far from the first to do that, but I think it fits into that tradition.

RF: There's an equivalent in the cynical tradition – e.g. Peter Sloterdijk's work most recently. This idea of putting some humor back into German idealism, I suppose. It made me think about some of the Silicon Valley kind of techno-solutionism which is so po-faced, as if they are the saviours of humanity and everyone else just has to shut up and fall into line. Give us all the money! Give us all the energy! Give us all the water! and we'll somehow save humanity, and not even the slightest hint of irony or humour ever, right? It's quite striking.

EC: One focus in the parrot paper was about the waste of these giant data sets, and the idea that if you create them there's bound to be bad stuff in there, all sorts of really horrible stuff in there, and it's kind of trapped in there. Also, there's the fact that it's all been stolen. Rob mentioned this idea of AI as a laundromat. It's like the AI companies are laundering the data.

I can't explain this as well as the former CEO of Google, Eric Schmidt. He was saying this to graduates in Stanford Business School:

> *"If TikTok is banned, each of you should tell a large language model, "Help me clone TikTok, capture all its users, steal all its music, add my personal preferences, write this program in 30 seconds, and publish it." Then wait an hour to see if it becomes popular. If not, make minor adjustments and try again.... while I obviously don't condone illegal music theft, I want to say that if your product becomes popular, you can hire a bunch of lawyers to sort everything out. If no one uses your product, don't worry — no one will care that you stole someone else's content. Don't tell anyone I said this. In short — copy, then use money to settle legal issues — that's how a lot of innovation in Silicon Valley actually operates."* (Liu 2024)

You couldn't make this stuff up...

RF: What do you think about the metaphor of AI as a laundromat?

EMB: Yeah, there's a lot in there. That Eric Schmidt quote is just horrific, *and* betrays a complete ignorance about how large language models work. No, you can't instruct it to do that for you.

It also betrays ignorance about what makes a social network valuable, right? It's the people and the community and the interactions, right? And then the like, "just do it and the lawyers will fix it". It is sort of the quiet part out loud, with the move fast and break things right? So that move fast and break things metaphor was supposed to be: "We're a startup. Go ahead and be bold as you're *coding* right? Don't be too timid." It was supposed to be, I think, literally about software code, but, in fact, that's the Silicon Valley ethos. And it's about social contracts and social systems. And disruption is just valued in Silicon Valley now with no thought to what the value of the systems is that are being broken through the disruption. So all of that is in the background. The AI as a laundromat metaphor sort of makes sense. Because if you take all of

the data and munge it together it stops being individual people's creative work and becomes a data set. And then that data set is the thing that they have.

I was invited to a presentation by the Department of Homeland Security about how they were implementing their part of the executive order on AI that came out of the Biden White House, and one of their headings was that they were concerned with intellectual property and by that they went out of their way to specifically say, it is not the intellectual property of the people whose data was scraped but rather the trained weights of the models that the companies had. That was the IP that needed protecting. So it's this idea that if you are smart enough at your engineering, you can take all this stuff that just becomes undifferentiated data and turn it into something valuable, I think, it's a kind of laundering.

RF: And if you make the pâté fine enough, then no one can ever tell where anything came from anyway so keep making a finer and finer pâté. Pâté is another metaphor. The context for asking this question is partly around open education, which is one of my research areas where people use open licenses, often by Creative Commons, to indicate how people can use their intellectual property for different things, you know, reuse, remix, and so on. There are some people who are saying: look, this GenAI stuff's great, because now we can just feed in all kinds of copyrighted stuff and out the other end, what comes out, is basically public domain by default, because it can't be copyrighted (because a person did not write it). Whereas other people are hand-wringing as follows: We put all this effort into creating an entirely separate system of intellectual property and open licensing, which now is meaningless, because no one cares anyway now about anything. People are taking as much stuff as they can, copyrighted or not, and they couldn't care less about what license you had for it. And so you've got some people in the open education space, trying to think, how can we change our licensing so that we can or can't allow this kind of training to happen? But people don't care. They don't care what license you've put on it. They're just taking it, anyway. And so this is where this kind of laundromat idea appeals to the idea of basically taking your dirty IP that you've stolen and washing it through a generative AI, and so that at the other end, you're free to do what you like with it.

EMB: The idea that GenAI is the future of open education resources is terrifying to me because GenAI means, you know, synthetic media that nobody has accountability for, and that seems anathema to what you'd be trying to do in open education, right? Open education is about – I'm speaking as an outsider – but I imagine it's about creating resources that other people can use, but that you, as the resource creator, stand behind. As in, I made this for this purpose and here's my sources. But then if you put it through... it's almost like it's almost as if the laundromat were full of blenders, right?

So, you know, you put in clothes that you have dirtied by stealing them, and out comes, you know shredded, you know, pulp, that someone then squeezes back into some shapes that may not be functional.

RF: Or it's like money laundering, right? You put in the dirty money at one end, and the clean money comes out the other.

EMB: Well, yes, except that the clean money is money, and it's functional as money, right?

So I want to point out, though, that in the stochastic parrots paper we missed the point about the data theft: it's not really in there, I don't think. We were talking about this drive to ever larger data sets, which takes you to a place where you can't feasibly work with them. If we can't feasibly know what's in your data set then you open up all these risks about what kinds of garbage is in there. And the other part of the story that's very clear now, is it also means that you basically have to steal it. And people are on record, such as Nick Clegg from Meta, saying asking artists for permission to use their work to train AI would "kill the industry" (The Verge 2025). OpenAI (2024) also argued along these lines in a submission to the UK House of Lords Communications and Digital Select Committee and Sam Altman has been saying "Yeah, if we had to pay for this, our business model would fall apart" (Milmo 2024). Which is ironic, because OpenAI are just losing lots of money anyway.

EC: There is a paper by Sfard (1998) on two metaphors for learning and the danger of choosing just one. It talks about the Frierian banking metaphor of education – where you just pour the information into students' heads – and then the participation metaphor. Sfvard says it's too

simplistic to pit one against the other but I think there is a risk of overusing the content/banking metaphor and or seeing content as king – especially in education.

You mentioned that you missed the extent of the data theft that was to come in your paper. What else did you find out that you didn't predict? Did you foresee the impact of AI on education?

EMB: Who knows the extent of what else we've missed? The paper was written in 2020 and time marches on. But the first thing was the combination of data, theft and labour exploitation. So it's not just that they're taking people's IP, but there's also terrible exploitative labour practices around data labelling that have since come to light, with respect to OpenAI in particular. Although there was already work in computational linguistics, looking at crowd work as problematic and a likely site for exploitation (Fort, Adda & Cohen 2011).

Anyway, that's one area that we didn't talk about. And then the other thing: that was I said we thought we were on thin ice when we were talking about synthetic text, right? When we named it the stochastic parrot we thought, well, okay, people aren't really going to want to do this. And, in fact the way that ChatGPT made it very accessible, like it was already starting a little bit with GPT-3, but with ChatGPT and that interface. You see it in education. You see it in legal areas, right? You see, not only lawyers, but also judges using ChatGPT. You see, all these proposals for psychotherapy over and over again, like people are incredibly enthusiastic about synthetic text, and we did not predict that. In fact, we were skeptical that people would really want to do much with it.

2. FRONTIERS & COLONIZATION

EC: I read two recent papers on bias, specifically in education and the biases of models – "The Psychosocial Impact of Gen AI Harms" (Vassel et al. 2024) and "Laissez-Faire Harms: Algorithmic Biases in Generative Language Model" (Shieh et al. 2024). They pointed out that they were harmful to the erasure of indigenous learners, harmful to black and Latinx learners and women, and biased and discriminated against them. And this is something that's huge in your [parrot] paper: bias. The bias issue is probably the biggest thing that you cover, and a question for you is – I'm going to give it a bad metaphor – it's like the bread and butter bias, right? It's the kind of bread and butter of this work. But it's also kind of boring. And people are like, oh, we've solved the bias? We're sure we've heard about this already? Surely it's done?

And you've been working on that. Your Mystery Hype Theatre podcast has been doing tireless work on this.

But the big AI companies are coming along. You're doing the bread-and-butter bias work. They're selling nothing-burgers and people are buying the nothing-burgers.

EMB: Yeah. So I think that, you know, the bias thing is interesting, because this was the site of most of the energy in natural language processing around societal impacts and so-called ethical NLP in like 2017 through 2020. And it continues.

So basically, people looked at what's called word vectors. If you take a large language model, you can pull out of it a representation of any given word that reflects how close it is in its textual distribution to other words and those word vectors reflect a bunch of things, sort of all in one combined representation. So they reflect things like grammar. Is this a noun or a verb? Is this plural or singular? They reflect things like lexical semantics. So *dog* and *cat* are going to have some similarities because they're both domesticated animals, and we're going to talk about them in some similar ways. But because this is all based on word distribution.

So there was a lot of energy. People thought they were doing good work because they were solving this problem. They're tackling this difficult issue of bias. But when the solutions were all technical solutions, they were sort of just staying in that tech solutionist space. So the bias is super important and documenting it is really important. But then people either stop, like, okay, it's been documented, or they try to come up with a tech solution for it.

To take a concrete example, you get things like Meg Mitchell points out that you never get the phrase, "man doctor", but you get "woman doctor" a lot. And this reflects a societal bias that doctors by default are men, which is hardly true anymore. Or rather , I should say, hardly true in the current moment, because there are historical moments where medicine was, I think, in

many cultures, actually a female profession. We ended up very male-dominated for a while in the West. But anyway, all of these biases are reflected in language, and that's a pernicious issue .

But you can bring up issues of bias in polite company much more easily compared to some of the other stuff. And so these word vectors represent this, and it was very quickly obvious that if you just ran these systems without doing anything about those biases, you're going to be reproducing the bias. And so you end up with a whole bunch of work that looks at measuring the bias around lots of different dimensions, and then looks at trying to mitigate the bias, using technical solutions. The thing to do when technical solutions aren't working because they aren't appropriate is widening the lens. Why are we using this technology in this way? Why have we gotten to this place where we have to collect so much data that not only is there going to be sort of run-of-the-mill biases in it, but we've sort of had to scrape the bottom of the barrel, and we've picked up these really awful parts of the dark web, so there's even worse stuff in it?

EC: Josh Tyrangiel (2024) of the Washington Post in February 2024 – I don't know if he feels egg on his face for saying this – he described Khanmigo as as a safe and accurate tutor, built atop ChatGPT, that works at the skill level of its users, never coughs up answers, is the best model we have or how to develop and implement AI for the public good. And it is also the first AI software that I'm excited for my kids to use.

He said it was safe, at the start, you know.

EMB: Yeah.

EC: But, it appears there are some really unsafe elements in what you say, and that there is work we have to do politically and socially together to combat this. It's not just a technical problem...

EMB: Yeah. And you know, safe for whom? Right? It may well be safe in some ways for kids who come from a lot of power and privilege, while being very unsafe for other kids. I still don't think it's safe even for kids with power and privilege, because they're still going to be getting nonsense. And they're still going to be taught that the synthetic media is a good way to access information, like that's all bad, like, just from the start.

But then there are different additional dangers depending on who you're talking about (Blodgett et al. 2020). But in the context of people suggesting that synthetic text extruding machines are a good information access system where, instead of going to a library and having a conversation with a librarian about how to navigate different sources, or instead of going to an old fashioned search engine and getting 10 blue links, and then exploring those links and sort of understanding how your question relates to them or doesn't, and maybe refining your question, you put your question in and you get an answer back!

That's got lots of problems (Shah & Bender 2022). One of the problems is that the answer looks more authoritative. So in that context, I have this example, where for a while, with Google snippets, so this is like, this isn't even the AI overview, it's just text that's been pulled out of the page: If you search "When did people come to America?" you got an answer about the first European colony right?

Well, what does that say about what people are? And at the same point in time, if you searched, when did *humans* come to America, you got an anthropological answer "About 33,000 years ago" or something, and of course, neither of those reflect traditional Native American views of since time immemorial. So there's a third thing that doesn't come out at all. But this is like this if you go to it.

And so you know, in this talk that I give, I have this example. Okay, imagine that you are the kid who's asked this question and you get back these answers. Well, what if that kid is non-Indigenous? What does that make them learn about their Indigenous peers?

And what about the Indigenous kid who's just been told: Oh, your people aren't people?

EC: So that's a metaphor, that generative AI is a colonizer (Gupta et al. 2024; Ferreira, Lemgruber & Cabrera 2023; Drumm et al. 2024). It's like a colonial force.

EMB: Yeah. There's an interesting deployment of that metaphor by Shoshana Zuboff in *The Age of Surveillance Capitalism* (2019), where she draws this extensive parallel between the European

colonizers and their land grabs in the New World and the tech companies, and the way that they laid claim to data and digital space. And I think that this is an interesting exploration of that similarity.

But I also think that we should be careful in using colonization as a metaphor (Tuck & Yang 2012) because I don't want to speak over Native Americans – in a North American context.

EC: I know what you mean, and it's very tricky for us in Ireland as we were a colonized country for hundreds of years, we suffered a famine and various things, but now we're rich white people with lots of money. We've got Google's European headquarters here and Amazon's. Lots of others too.

So there's only so long we can talk about being an oppressed people because we're clearly not anymore. And we have a lot of problems in our own society, including an indigenous minority population, the Irish Travelers who have the worst life outcomes of any group in Ireland.

RF: I've been pulled up on this before. Using "colonization" in the [Jürgen] Habermasian sense and this is the idea of taking over the world that people live in with kind of capitalist forces essentially. When I tried to use this to talk about open educational spaces, and talking about basically recolonizing educational spaces with open educational resources, some people were like, "we don't like this language that you're using." I wasn't even thinking about new world colonization. I was thinking about this theory in sociology.

EMB: Yeah. So there's also related language in the description of large language models that are sometimes called "frontier models", which has similar problems. And this, I think, again, speaks specifically to the colonization of North America. Where there's this notion of like, there was, there was the initial land grab by European powers. And then, after US independence there was American colonization further West and the frontier was this like unexplored virgin territory. And you have all of the mythology of the American West as being basically empty before the European settlers came, which is false. When they talk about large language models as frontier models, I think they're talking about, not this history of theft and destruction and murder. They're talking about the romanticized view of it as the exciting, like, the leading edge of development. And so it's very...

EC: ... next generation?

EMB: Exactly! Star Trek.

EMB: Yeah. I don't know if we want to stick Star Trek with this. But yeah, exactly. It's this, this like, the excitement of exploration. And it is so unreflective to just adopt that term in Silicon Valley. If you think about where they are and the history of that place. It is sort of appalling that they would just use the term frontier models, and not sort of relate to the history of what that means.

RF: Maybe it's not an oversight, right? Maybe it's intended.

EMB: Maybe. I would think it's more likely just utter privilege-derived cluelessness, actually and the inability to imagine why it might be problematic or hurtful.[1]

KL: I would like to talk a little bit more about this issue in an educational context. There are two things that I was wondering about this metaphor that are very common in educational settings: that AI is going to be an assistant, and obviously a tutor and helper and friend, the positive connotation that we have toward AI[2] – I try to refute that. And then people's reaction is usually: were human teachers any better? It's a random distribution of words, but still, AI or ChatGPT will answer students, whereas human nature wouldn't necessarily do anything. So what is better and worse? People got really haunted by that.

It has, unfortunately, had this impact upon teachers, and how we see teachers, and how people used to see teachers as human and not being capable. So I think that it's a kind of dilemma, where as soon as we raise AI up, then we automatically really devalue human teachers and

1 For an interesting essay on this topic see Barbrook, R and Cameron, A. 1996. 'The Californian ideology'. *Science as Culture*, *6*(1), 44–72.

2 This type of 'rosy' scenario is explored in the short story 'The two sparrows', in Lee, K-f and Chen, Q. 2021. *AI 2041: Ten visions for our future*. New York: Crown Currency.

human agency and our own capacity. I just wanted to talk about your view on that. And you know it's like this post-human world. So what shall we do about that?

EMB: Yeah, yeah, it's frustrating. I think that it's an argument I see in many, many places where you say: well, but a human would. And I was just reading this absolutely terrible journal article on AI and ethics, arguing that we should allow large language models as co-authors of papers and half of the arguments were well, but people also do x, y and z. It's a false equivalence. It's when a teacher gives an incorrect answer that you can sort of say, Well, why is that? Where did that come from? What do we do about that? And the answers to that are going to be very different from "ChatGPT extruded a sequence of tokens that led the student to believe something incorrect." Okay, well, why is that? What do we do about it? It's very different. But also, then it's inherently dehumanizing (as you point out) to say, well, you know people are just as bad as this. I think it's also illuminating to look into sort of what led us to this moment. So if a teacher can't answer the student, is that because there's not enough resources being put into education, and so teachers only have so much time in the day? Don't students have more needs than the teachers can meet?

This is something that comes up so often on the podcast where we say, okay, just because you've identified a problem doesn't mean a large language model is a solution to it, right? Teachers are under supported. We need more resources in education. True, this is a problem. Okay, let's give the kids a fake text! That is not a solution. And the argument of "Well, it's better than nothing." Well, why should we settle for just better than nothing? And also, is it really better than nothing? You have a kid who has a question and a teacher who doesn't have time to answer. Is the better outcome that the kid is told: "Oh, just take the word of this machine for it," and the machine is going to be outputting text that looks confident versus "wait" versus "ask your friends" versus "try the library" versus, "think about it yourself"? If a teacher doesn't know the answer, that can be a great learning moment. This is something that I tell teaching assistants at my university. I say "I have three magic words for you that you can use to answer any question and the magic words are: 'I don't know'." And then the magic is what happens next. "I don't know. Let's think about how we could find out. Does anybody else here know? Let's go look it up." All of that is actually modeling learning in a way that ChatGPT, coming out with something that may or may not be right, absolutely isn't. I have strong feelings about this.

KL: Yeah, I like it a lot! And another question I was having, kind of if we move back to colonizers. Now, it's been very interesting. I've moved back to Korea. We were colonized heavily by Japan, and we're still very sad about it. But anyway, for us, a lot of people here say, we tend to use this metaphor that all this translation that's offered by AI, see it as opening and the entrance to the world, that for the first time we have access to the entire knowledge, and then we can be on the same line to compete with Westerners. So actually, the people who are colonized are not necessarily buying into the argument that this is doing something bad to us.

They're actually very happy and they're excited about it. So it's kind of interesting that the former colonizers, some of them, or like a small number of them, are much more critical. They have critical awareness of what they have done to people, but from our side of those people we are not really seeing it. So at the beginning of that I thought, like people are too naive but actually I could see a lot of good cases coming out of it. People are enjoying these possibilities. They can get into somewhere that they were not. The door was locked previously, but it's like open if they can push it. So now I'm kind of seeing I'm not really sure what to think about it.

EMB: I think we want to separate out machine translation from other things that get called AI. Machine translation definitely has downsides, but I think it is a reasonably well-defined technology. In contrast, part of the problem with the large language model AI systems is that they are what Timnit Gebru calls "everything machines". And you can't actually evaluate an "everything machine". You can't tell how well it does at everything, because you can't possibly test that. Machine translation is something that you can test. You're going to want to test it for the specific language pair you care about. You're going to want to test it for the kinds of documents you're looking at. I think there's some really important design work that has yet to be done in machine translation to make it clear to the user how much uncertainty there is.

Machine translation can be very good for figuring out what a text is about, generally. But historically, machine translation has been very poor at translating negation in particular. So you can end up with things that mean the opposite of what you want, and the machine translation

is also really bad with pronouns across languages, and that's something that going from English or French or German into Korean probably wouldn't be too bad. I think Korean tends not to use pronouns right? The machine translation system is going to output a pronoun, because English requires that the English language model says there should be a pronoun here, but there's no machine "translation". It is not a situation where the system has understood what was said in English, and then written it out in Korean, or vice versa, right? It's a statistical mapping between the languages. I think there's a lot of value to be had in automatic translation and automatic transcription, but it needs to be packaged in a way that people can see the uncertainty. I think the ideal use case for machine translation is figuring out what's the document that you then want to translate and that can speed things up a lot. For if you're talking about a situation where it feels like all of this discourse is happening in an inaccessible language, I don't have time to carefully read it all. But I can use machine translation to go find the ones that I really care about, and then do that more carefully.

3. SOLUTIONISM, POSITIONALITY & VOICE

RF: Some people feel in education that AI is a kind of solution in search of a problem to solve right? Because it's not really a specific solution that's needed in education. It's more that there are all these possible uses. People are trying to flush out: where does it work? Where is the actual gain to be had? But in the meantime we're going to get it added to everything.

EMB: And baked into Microsoft office, and, you know, all the rest of it.

RF: So there's little escape from it, even though it's not actually there to solve a problem as such, or at least it solves someone's problem. But it's not the educator necessarily.

EMB: And when you say a solution in search of a problem, of course, solutions don't search right. It's the people selling the solution who are searching for the problem. They're searching to create it. And I think it's really interesting to contrast the people who are incorporating so-called AI into Microsoft Office i.e. Microsoft, with people, say in Korea [in Kyungmee's earlier example] who are interested in getting access to things that are written in other languages and thinking about what are the motivations? Who has power in these situations? Who has control over how the technology is actually developed? Who has control over how they're using it? This isn't to say that people seeking to use machine translation are powerless. They have a lot of choices, but I think that there is more that could be done to make the tools transparent enough that people could really be effective and creative in how they use them. And that's the kind of solution-seeking that I think would be empowering rather than the problem-seeking that you're referring to.

RF: It eventually becomes quite mundane! Something like having a phone in your pocket, a computer in your pocket that can do all these things is kind of just so mundane now as whereas 20 years ago that was a crazy idea, right? And I just read about a protest regarding people using calculators in schools in the 1960s because they didn't want anyone to become dependent on a machine for calculation. I remember when I first started learning word processing and stuff like that in the late 1990s. And some people didn't like it because you'd become too reliant on the spell-checker, and you'd stop thinking for yourself about how to spell things and stuff like that. But we think of that as a bit kind of twee almost. Now, you know the idea of like oh, you're too dependent. You can't spell without Microsoft Word or without Clippy popping up, or something. Do you think it will be something similar with this kind of technology in the future, where it's just so mundane and so everyday, or is there something else happening?

EMB: So the difference that I see between calculators and spell checkers, on the one hand, and large language models, on the other, is that calculators and spell checkers have particular tasks that they do, and they do them correctly. Whereas ChatGPT is this "everything machine" that does nothing well. And so we could go down that path. But that'd be very dystopian. That'd be very glass empty, as Eamon was saying...

EC: Yeah, it's snake oil. Martin Weller does a lot of work with metaphors in educational technology, and he wrote recently about AI as the insecticide DDT (Weller 2024), saying that it's like a poison that gets into the environment. DDT did certain things kind of well, it worked in

one context, but it was just applied everywhere, and people didn't know the harms, and that seems like somewhere we could be heading to. I don't know what you think about that?

EMB: There's a link that I found that we're probably going to talk about on our next podcast episode (Bender & Hana 2024), where somebody is unironically saying, "AI is the new plastic." What they're trying to say with that metaphor is, it's useful in so many ways it can be shaped into anything, and if you have any awareness of environmental issues, right? It's exactly this thing. If it's actually going to get everywhere, it's going to be polluting (in this case our information ecosystem) in ways that it's really really hard to remove. With, you know, forever chemicals, and all this kind of stuff, why would you say AI is a new plastic as a positive metaphor? It's actually a pretty effective metaphor, but not for the reason they think.

EC: It's been called the microplastic of the Web. It's also been called plastic for your cognitive agency. Are there any metaphors from your new book (Bender & Hanna 2025) that you can share?

EMB: So you know, I was thinking about that. And certainly the title of the book actually is a metaphor, right? We talk about "The AI Con". And that is not so much a metaphor for AI as a thing in the world, but a metaphor for the actions of the companies selling AI and it's maybe not the most creative or visually striking metaphor. It's more like we're saying, "this is a con," but I think that that might qualify as a metaphor. The basic idea is that we're being sold a bill of goods and people are telling us that these systems either already solve all these problems or are about to. There's all kinds of citations to the future. And they're telling us that not just as a prank or as a joke, but to make money. And so that's sort of, in a sense, the central metaphor of the book.

KL: Emily, do you think that metaphors can be neutral?

EMB: Can metaphors be neutral? Can anything be neutral? I mean, I think that anytime someone is developing a metaphor they're doing it to explain something. And if you're explaining something, you're doing it from some point of view. So, I think there's probably some metaphors that are much more pointed, or are going to come out as politically much more to one end or the other. I suspect that metaphors always, just like any kind of science or scholarship, have a point of view.

KL: Having said that, I was wondering because the special issue we're trying to create right now has political intention in that I think all of us are working in a space where we just wanted to kind of resist rules and norms in terms of how to write academically. And I think, yeah, we're sick and tired. So we're kind of trying to explore different ways in which we can be artistic and more imaginative and do new things. This special issue from my perspective has two different purposes. Firstly, we wanted to really talk about AI. But secondly, we want to tell people that it's okay to be, you know, playful, creative, let's just use metaphors and things like that. So, it would be really interesting to hear your opinions, and more like encouragement, on how to do those things.

EMB: Absolutely! There's a new journal out of Duke University Press, but run by Rutgers called *Critical AI* (2025), and they're also very much sort of humanities-based and pulling on different forms of expression. So you might enjoy that as sort of a kindred project in terms of academic expression. Metaphors always have a point of view, but so does all academic writing. There's this claim to an objective viewpoint, and it's false. It is also definitely a power move, in that if you look across the sciences – and I'm speaking mostly from my own location here in the United States of America, but I suspect you'll find this elsewhere – there's this notion that the people who experience the least marginalization are the best able to present themselves as having objectivity, and therefore have the easiest time being taken seriously as scholars. I think the thing to embrace is the idea that there is no neutral position, and that we do our best scholarship when we own our positionality (Haraway 1988). It's this idea that there is no view from nowhere, and that anybody who's espousing that is actually making a very strong political claim, but hiding it, and that makes it stronger.

The other thing that I want to say is that – speaking very much as a linguist – a lot of that plays out in English academic writing with use of passive voice, right? Sort of hiding agency, and, you know, refusing to use first person pronouns in academic writing, and also in writing that is very abstract and sort of harder to understand. There's a sense that if what I write is difficult for

people to read, then they'll assume that I'm smart and sort of accord me more credence and and more status. I think honestly, the hardest thing to do is to write clearly about new ideas, and I think that metaphor and playfulness, and putting yourself as the scholar into the writing, are things that make it clearer.

RF: It's interesting just to hear you talk about this because I'm thinking about philosophy, where a lot of emphasis is placed on trying to arrive at this clear writing style but the way that a lot of philosophers write is actually quite exclusionary because unless you've got a background, and understand the text they're referring to, and the ways the arguments are working, it's really hard to get a handle on anything.

In philosophy we have a long history of metaphors, which goes back to Plato at least, trying to convey these things in a more accessible way. They can be more inclusive, but also exclusionary at the same time, because it depends on what your frame of reference is in the first place, right? If I start using a metaphor, and you don't even really understand the thing that I'm using, as a metaphor, it has no weight, you know. It doesn't really cut through in the way that it's supposed to. So I guess it's inclusion and exclusion, at the same time depending on how your chosen discourse approach, whether it's metaphors or writing philosophy papers, or whatever it is. There's always going to be some dynamic there like that.

EMB: Yeah, yeah. And I also wanted to react to this idea of pushing back against norms of academic communication. Because I think that there's some that are really worth pushing back against – the view from nowhere, passive voice, all that stuff. But then there's some that I think are really important to uphold, like citational practice. And you know, making clear where ideas are coming from and giving credit and situating work with respect to other things. And so that's something where I think there's value in tradition, and it probably needs some improvement rather than throwing it away.

EC: Yes the AI "voice" that is doing the communicating is interesting too. The philosopher Liz Jackson had a great piece out recently called "AI's Manliness Problem" (Jackson 2024).

EMB: Yeah, People refer to ChatGPT as mansplaining-as-a-service which I like (Jackson et al. 2025). But then, like you know, double edge, as always it makes it sound thoughtless rather than mindless. I don't think mansplainers lack internal life right?

EC: So reading the parrot paper, Emily – I am from Ireland, so I'm compelled to talk about Palestine. We identify with the Palestinian cause. That is what a lot of Irish people do. You mentioned actually Palestinian bias in the paper. This idea of minoritised people and bias, and so on. But also, what resonated with me and hit me quite hard when I was reading your paper was that (and I guess this applies to algorithms, or any sort of general recommender systems or social media in general) is that the big, angry kind of thing like "the protest that turns violent", that's the one that gets highlighted by the algorithm, by the language model. Whereas peaceful protests don't get attention. They don't become prominent in the outputs, and they get kind of lost in the wash. This was really depressing to me, and I was thinking, "this is horrible!" But then I actually had a thought that maybe that's a good thing in some sense: that the peace is beyond the algorithm? It can't take peace – whatever peace we can find, ourselves or together, that's something beyond the algorithm.

EMB: Look at you being hopeful. The glass is filling up, Eamon.

If you're training from collections of data, then you're beholden to, or you're dependent on what went into that data in the first place. The way our news media works is that the attention is drawn to violence and not to peace. If peace shows up in the news as a hopeful future thing, I think, and never a present thing. (I'm making a generalization there, but that's a guess.)

So that's all that can come out of the language models, because that's what's in the training data. If we say it's a good thing that what doesn't show up in the training data is ours to keep, and I think it depends a bit on how we operationalize what we're doing with the machines, right? My bias is, "let's do as little as possible with the machines, and then it's not going to matter as much." However, there may well be cases where that doesn't go the way that I want. So, the fact that there are certain things that remain outside, because they're not written about so much, is, you know, perhaps a kind of solace in the storm.

4. BULLSHIT & ENSHITTIFICATION

RF: Maybe one metaphor we didn't cover was AI is bullshit – Harry Frankfurt's bullshit (Frankfurt 2005; Hicks, Humphries & Slater 2024; Costello 2024).

EMB: I think this one can still be anthropomorphizing, because (at least the way it's commonly understood) if you're bullshitting you are speaking without caring whether or not it's true. And ChatGTP can't even not care, because there's nothing there to withhold care. So I think it's a little bit too anthropomorphizing, although it is better than hallucination. This is why I tend to talk about *synthetic text extruding machines*, which is again, maybe not super metaphorical but I do like the phrase and there are some metaphors embedded in the word extrude.

EC: And indeed it's pulling on ideas like plastic, or in my childhood Play-Doh was a thing. Did any of you experience that?

EMB: I did write an essay in high school about how Plato and Play-Doh are not that different. And it was written as a dialogue with my then three-year-old brother.

RF: Fair enough!

EMB: Yeah, so this is the toy, Play-Doh. One of the versions of it we could get in the USA came with a bunch of plastic accessories, including something that would be like a pasta making machine. So you stuff the Play-Doh in on one side, and then cranked it, and out came these tubes of this inedible stuff. So that's the kind of stuff that I want to evoke with "synthetic text extruding machine". So what comes out is synthetic rather than authentic, and it is displeasing in the way that it comes out. But bullshit is better than hallucination, for sure.

EC: Neil Selwyn had a paper on how edtech is full of bullshit (Selwyn 2015) where the metaphor works at the level of the discourse too. We seem to be going through a scatological moment as well, because with this idea of enshittification (Doctorow 2022) – we've never seen it happen faster than with the AI tools.

EMB: Right.

EC: I have a horrible sense of everything, sort of speeded up now, you know, everything's getting crap so fast.

EMB: It is, and so many people seem to want to be on the bandwagon rather than seeing it for what it is, which is frustrating. I think that's part of the speed, but also part of the speed is the amount of money behind it. I had the opportunity to speak with a bunch of deans from universities along the west coast of the USA and Canada about so-called AI, in March of 2024. And I got to tell them that this is becoming your problem now, not because there's been some major technological breakthrough which is how the media sometimes portrays it. It's becoming the problem of deans and other educators, because there's people trying to sell this, and they have a lot of money behind them pushing those sales. I think that's an important thing to always keep an eye on. But when you say that the edtech discourse is full of bullshit, I'd believe that.

RF: In a way, you've always got that pressure in the edtech space, because there's always new products coming to market. So there's always going to be people trying to create those markets. There's always going to be people trying to sell to the purchasers who have power at scale. They need to always be creating those new markets and be the first one to benefit from it, right? So that's what they're all fighting over.

EMB: I think because there's a concentration of resources that people want to get access to and it makes sense that we should concentrate resources and use them to provide education in a you know, not family level, but larger community level fashion. I think that makes sense. We were talking before about needing more resources there. We need teachers to be supported better. But once you have those resources concentrated, they become an attractive force. And I think there's a similar thing going on with data. So if you think about the data that's generated by our individual interactions on the web, each one of those on its own is not super valuable. But if someone manages to collect a concentration of it, then it becomes this valuable and attractive thing, and needs protecting. But oftentimes the better choice is to not collect it in the first place.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Emily M. Bender orcid.org/0000-0001-5384-6227
University of Washington, US

Eamon Costello orcid.org/0000-0002-2775-6006
Dublin City University, IE

Kyungmee Lee orcid.org/0000-0002-9580-9026
Seoul National University, KR

Robert Farrow orcid.org/0000-0002-7625-8396
The Open University, UK

Giselle Ferreira orcid.org/0000-0002-8498-5390
Pontifícia Universidade Católica do Rio de Janeiro, BR

REFERENCES

American Dialect Society. 5 January 2024. '2023 Word of the Year is 'Enshittification'". *American Dialect Society* [online]. Available at https://americandialect.org/wp-content/uploads/2024/01/2023-Word-of-the-Year-PRESS-RELEASE.pdf (Last accessed 22 July 2025).

BBC News. 20 February 2021. 'Margaret Mitchell: Google fires AI ethics founder'. *BBC News* [online]. Available at https://www.bbc.com/news/technology-56135817 (Last accessed 10 July 2025).

Bender, EM. 2024. 'Resisting dehumanization in the age of "AI"'. *Current Directions in Psychological Science*, 33(2): 114–120. DOI: https://doi.org/10.1177/09637214231217286

Bender, EM, Gebru, T, McMillan-Major, A and **Shmitchell, S.** 2021. 'On the dangers of stochastic parrots: Can language models be too big?' *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. DOI: https://doi.org/10.1145/3442188.3445922

Bender, EMH and **Hanna, A.** 9 December 2024. *Episode 47: Hell is o*ther People's *AI hype*. AI Mystery Hype Theatre. Available at https://www.buzzsprout.com/2126417/episodes/16353749-episode-47-hell-is-other-people-s-ai-hype-december-9-2024 (Last accessed 10 July 2025).

Bender, EMH and **Hanna, A.** 2025. *The AI con: How to fight big tech's hype and create the future we want*. New York: HarperCollins.

Blodgett, SL, Barocas, S, Daumé III, H and **Wallach, H.** 2020. 'Language (technology) is power: A critical survey of "bias" in NLP'. *arXiv preprint* arXiv:2005.14050. DOI: https://doi.org/10.18653/v1/2020.acl-main.485

Costello, E. 2024. 'ChatGPT and the educational AI chatter: Full of bullshit or trying to tell us something?' *Postdigital Science and Education*, 6(2): 425–430. DOI: https://doi.org/10.1007/s42438-023-00398-5

Critical, AI. 2025. Available at https://sites.rutgers.edu/critical-ai/ (Last accessed 22 July 2025).

Doctorow, C. 15 November 2022. 'Social quitting'. *Locus* [online]. https://locusmag.com/2023/01/commentary-cory-doctorow-social-quitting/ (Last accessed 23 October 2023).

Drumm, L, Beetham, H, Cronin, C and **McIlwhan, R.** 2024. 'Generating AI alternatives: Collaborating and creating intersections'. *Proceedings of the International Conference on Networked Learning,* 14(1). DOI: https://doi.org/10.54337/nlc.v14i1.8185

Ferreira, GMdS, Lemgruber, MS and **Cabrera, TL.** 2023. 'From didachography to AI: Metaphors teaching is automated by'. *Journal of Interactive Media in Education*, 2023(1). DOI: https://doi.org/10.5334/jime.798

Fort, K, Adda, G and **Cohen, KB.** 2011. 'Last words: Amazon Mechanical Turk: Gold mine or coal mine?'. *Computational Linguistics*, 37(2): 413–420. https://aclanthology.org/J11-2010/ (Last accessed: 22 July 2025). DOI: https://doi.org/10.1162/COLI_a_00057

Frankfurt, HG. 2005. *On bullshit*. Princeton, NJ: Princeton University Press.

Gupta, A, Atef, Y, Mills, A and **Bali, M.** 2024. 'Assistant, parrot, or colonizing loudspeaker? ChatGPT metaphors for developing critical AI Literacies'. *Open Praxis*, 16(1): 37–53. DOI: https://doi.org/10.55982/openpraxis.16.1.631

Haraway, D. 1988. 'Situated knowledges: The science question in feminism and the privilege of partial perspective'. *Feminist Studies*, 14(3): 575–599. http://www.jstor.org/stable/3178066. DOI: https://doi.org/10.2307/3178066

Hicks, MT, Humphries, J and **Slater, J.** 2024. 'ChatGPT is bullshit'. *Ethics and Information Technology*, 26(2): 38. DOI: https://doi.org/10.1007/s10676-024-09775-5

Jackson, L. 2024. 'The manliness of artificial intelligence'. *Educational Philosophy and Theory*, 57(7): 645–649. DOI: https://doi.org/10.1080/00131857.2024.2409739

Jackson, L, Sidorkin, AM, Jandríc, P, Costello, E, Heybach, JA, Greenhalgh-Spencer, H, Hytten, K, Gourlay, L, Buchanan, R and **Tesar, M.** 2025. 'The voice of artificial intelligence: Philosophical and educational reflections'. *Educational Philosophy and Theory*, 57(7): 650–661. DOI: https://doi.org/10.1080/00131857.2025.2481588

Jandrić, P. 2024. 'The creation of myths as an educational strategy and emancipatory opportunity'. *Beijing International Review of Education*, 6(3–4): 191–207. https://journals.sagepub.com/doi/10.1163/25902539-06030002

Liu, W. 18 August 2024. 'Notes on Eric Schmidt's AI talk at Stanford'. *Medium* [online]. https://medium.com/@weikang_liu/notes-on-eric-schmidts-ai-talk-at-stanford-72432c848fde (Last accessed 15 July 2025).

Milmo, D. 8 January 2024. "Impossible' to create AI tools like ChatGPT without copyrighted material, OpenAI says'. *The Guardian*. https://www.theguardian.com/technology/2024/jan/08/ai-tools-chatgpt-copyrighted-material-opena (Last accessed15 July 2025).

OpenAI. 2024. *OpenAI—written evidence (LLM0113)*. House of Lords Communications and Digital Select Committee inquiry: Large language models. https://committees.parliament.uk/writtenevidence/126981/pdf/ (Last accessed 15 July 2025).

Selwyn, N. 2015. 'Minding our language: Why education and technology is full of bullshit … and what might be done about it'. *Learning, Media and Technology*, 41(3): 437–443. DOI: https://doi.org/10.1080/17439884.2015.1012523

Sfard, A. 1998. 'On two metaphors for learning and the dangers of choosing just one'. *Educational Researcher*, 27(2): 4–13. DOI: https://doi.org/10.3102/0013189X027002004

Shah, C and **Bender, EM.** 2022. 'Situating search'. *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, pp. 221–232. DOI: https://doi.org/10.1145/3498366.3505816

Shieh, E, Vassel, FM, Sugimoto, C and **Monroe-White, T.** 2024. 'Laissez-faire harms: Algorithmic biases in generative language models'. *arXiv preprint* arXiv:2404.07475. https://doi.org/10.48550/arXiv.2404.07475.

The Guardian. 19 February 2021. 'Google fires Margaret Mitchell, another top researcher on its AI ethics team'. *The Guardian*. https://www.theguardian.com/technology/2021/feb/19/google-fires-margaret-mitchell-ai-ethics-team (Last accessed 10 July 2025).

The Verge. 26 May 2025. 'Nick Clegg says asking artists for use permission would "kill" the AI industry'. *The Verge*. https://www.theverge.com/news/674366/nick-clegg-uk-ai-artists-policy-letter (Last accessed 15 July 2025).

Tuck, E and **Yang, KW.** 2012. 'Decolonization is not a metaphor'. *Decolonization: Indigeneity, Education & Society*, 1(1).

Tyrangiel, J. 22 February 2024. 'An "education legend" has created an AI that will change your mind about AI'. *Washington Post*. https://www.washingtonpost.com/opinions/2024/02/22/artificial-intelligence-sal-khan/ (Last accessed 15 July 2025).

Vassel, FM, Shieh, E, Sugimoto, CR and **Monroe-White, T.** 2024. 'The psychosocial impacts of generative AI harms'. *Proceedings of the AAAI Symposium Series*, 3(1): 440–447. DOI: https://doi.org/10.1609/aaaiss.v3i1.31251

Weller, M. 27 September 2024. Behold the new wonder chemical. *EdTechie* [online]. https://blog.edtechie.net/books/behold-the-new-wonder-chemical/ (Last accessed 15 October 2024).

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

TO CITE THIS ARTICLE:
Bender, EM, Costello, E, Lee, K, Farrow, R and Ferreira, G. 2025. Unsafe AI for Education: A Conversation on Stochastic Parrots and Other Learning Metaphors ⚠. *Journal of Interactive Media in Education*, 2025(1): 10, pp. 1–16. DOI: https://doi.org/10.5334/jime.1079

Submitted: 25 July 2025
Accepted: 05 August 2025
Published: 26 August 2025

Journal of Interactive Media in Education is a peer-reviewed open access journal published by Ubiquity Press.