The Looming AI Winter: Dewey & Dreyfus on Understanding, Expertise, and "Artifactual" Intelligence

In the spring of 1920, the illustrious British philosopher Bertrand Russell fell ill while on a lecture tour in China. Reports of his untimely demise quickly spread across world media outlets, with dozens of obituaries appearing in the following days – each commemorating aspects of the public figures' life. Months later, when Russell – still very much alive and fully recovered – returned to Britain, he invoked the words of Mark Twain as he quipped about the exaggerated reports of his recent demise.[1]

One century later there's another grossly exaggerated rumor about a famous death making the rounds in the dark corners of online forums and chatrooms. Since early 2020, proponents of "Internet Death Theory" (as it's come to be known) have argued that the internet "died" somewhere around late 2017 or early 2018 and is now "almost completely devoid of people."[2] Although this conspiracy is patently ridiculous in its details, the concern that motivates it is a legitimate one. After all, much of the internet IS fake.

Of course, for anyone who's ever been "cat-fished," duped by a "sock-puppet," or taken in by the pretenses of an "astroturf" campaign, this isn't exactly news. The evolution of the internet from Web 1.0 - "Hey, have you checked out this email thing?" ... to Web 3.0 - "Hey, my toaster just messaged my watch that breakfast is ready." is really a story of how a network of static information portals became an interactive, semantic network of user generated data - and, ipso facto, a hot mess of dissimulation and machination. Under the "big data" in our "Internet of Things"—as Web 3.0 is sometimes called—it's increasingly difficult to discern what's real from what's a simulation. One reason for this is the sheer volume of machine-generated data.

In its 8th annual "Bad Bot Report," the cyber-security firm Imperva concluded that 40.8% of all web traffic requests made in 2020 were generated by non-human users.[3] Alarmingly, the percentage of malicious bot traffic is at an all time high. And that is one reason why cyber-security specialists are desperately seeking a way to monitor and block bad bot requests. Because there's just too much data for human moderators to police, most sites employ algorithms to spot (and block) bot-generated traffic. But, there's the rub. Whenever there is more bot generated traffic than human traffic, there's a risk such algorithms, which rely on machine-learning, will begin to flag human-generated traffic as anomalous. Max Read, writing for New York Magazine, has dubbed this moment - when every REAL person on the Internet is automatically deemed to be fake - "the Inversion." He explains where he got the name:

> For a period of time in 2013... a full half of YouTube traffic was 'bots masquerading as people,' a portion so high that employees feared an inflection point after which YouTube's systems for detecting fraudulent traffic would begin to regard bot traffic

---

[1] What Twain actually said was: "I can understand perfectly how the report of my illness got about, I have even heard on good authority that I was dead. James Ross Clemens, a cousin of mine, was seriously ill two or three weeks ago in London, but is well now. The report of my illness grew out of his illness. The report of my death was an exaggeration." The account offered by Twain's biographer Albert Bielow Paine has given us the popular misquotation: "Just say the report of my death has been grossly exaggerated."

[2] Quoted in Tiffany, Kaitlyn. "Maybe You Missed It, but the Internet 'Died' Five Years Ago," Atlantic Monthly (August 31, 2021). https://www.theatlantic.com/technology/archive/2021/08/dead-internet-theory-wrong-but-feels-true/619937/?utm_source=copy-link&utm_medium=social&utm_campaign=share [Accessed 5/15/2022]

[3] https://www.imperva.com/resources/resource-library/reports/bad-bot-report/?utm_source=press&utm_medium=pr

as real and human traffic as fake... They called this hypothetical event 'the Inversion.' (Read 2018)

But, as Read suggests, this is not just an issue for some algorithms overreaching. Rather, we run the risk of even real, human internet-users dismissing one another as mere bots. In some ways, this Inversion has already happened. One benign example occurred in late 2018 when some were calling for a boycott against Netflix for allegedly astro-turfing a meme campaign about the film *Bird Box*. However, it turned out the memes had been created by actual fans of the film, as implausible as that sounds!

Of course, there are much less benign examples in the political sphere, which underscores the importance of addressing these issues, and the COVID-19 pandemic has demonstrated how our virtual world can be used toward very real benefit. So, where does this leave us? We obviously can't scrap the whole thing. Read aptly sums up our phenomenological predicament:

> The "fakeness" of the post-Inversion internet is less a calculable falsehood and more a particular quality of experience — the uncanny sense that what you encounter online is not "real" but is also undeniably not "fake," and indeed may be both at once, or in succession, as you turn it over in your head. (Read 2018)

But, in the face of such duality, perhaps "turning it over in our heads" is precisely what's needed. A good **Netizen** (citizen of the Net) needs to practice mindfulness and grace… all while maintaining a healthy dose skepticism. That's a bit of a philosophical tightrope… but it CAN be done. In what follows I'll discuss how, by leaning on the philosophical ideas of John Dewey and Hubert Dreyfus.

One example of that uncanny, real/fake duality Max Read has identified might be found on the Instagram profile of Miquela Sousa, AKA Lil' Miquela, a 19-year-old model and influencer—who also claims to be an artificially intelligent "bot."

First appearing on Instagram in 2016, Miquela's account featured photo shoots with celebrities, SoundCloud tracks, and "woke" political activism. Though speculation about her "realness" existed from the outset, she nevertheless gained a following in the millions, eventually being "outed" as a bot in 2018 during a feud with another Instagram influencer — who also claimed to be an AI bot. Surprisingly, this only caused Miquela's star to rise higher, *as she insisted that she had no idea she was not a real person*. Followers were riveted by her story and lined up behind their keyboards to voice their support, especially as her posts began to express a longing for liberation from her creators at Brud — a company purported to specialize in AI research.

Since then, Miquela has embraced her AI identity and seemingly reconciled with Brud. She's also featured in video ads for Calvin Klein with "REAL" influencers like Bella Hadid, attended Coachella as a host for YouTube Music, and joined Samsung's Galaxy Team to pitch their line of smart phones. Of course, the claims regarding Miquela's sentience as an AI bot aren't actually real. It's all part of a new genre of fictional story-telling that uses social media platforms to build narratives in real time. Miquela's handlers at Brud have built a fairly elaborate backstory and fictional universe around her, generating millions of adoring fans. As of spring 2022, Miquela has accumulated over 3 million Instagram followers. As one put it shortly after the true nature of her

identity was revealed: "I know this is crazy but I believe you. Even though you are a robot physically, everything else [about you] is human."

After all, Miquela has donated and/or raised hundreds of thousands of dollars toward LGBTQ+ rights and Black Lives Matter awareness. Emilia Petrarca of New York Magazine was one of the first to write about the Lil' Miquela phenomenon. In her words:

> Lil Miquela... holds up a mirror to the ways in which technology has morphed our own constructions of self. We don't yet live in a world where realistic-looking fake humans roam the streets, but in the meantime, technology has transformed us into fake-looking real humans...
>
> ...We spend so much time pretending that hip parties and cool people are an organic part of our lives — that we aren't curating the narratives we put out in the world. What a perverse relief it would be to confess that everything is fake! (Petrarca, 2018)

Truly, this is the most dizzying sort of Inversion, where people we <u>know</u> are fake… make those of us who are <u>real</u>… <u>feel</u> like total phonies - after all, how many of us have raised that kind of money or awareness for the causes we claim to support?

Although it's been several years since Max Read first noticed the Inversion and since Lil' Miquela first appeared on Instagram, philosophers have been slow to respond. That may have something to do with the way philosophers tend to view such phenomena — *i.e.* through the fairly narrow lens of <u>analytic philosophy of mind</u>.

Put simply, when philosophers of mind encounter claims about AI, they tend to focus on whether or not the machine in question can thoroughly convince others that it's thinking… as the the famous "<u>Turing Test</u>" first explored in the 1950s. But, considering that Lil' Miquela is REALLY the product of a team of human creators… pretending to be an AI bot… pretending to be a real human, such academic questions regarding "<u>strong AI</u>" are at best moot, and at worst completely antiquated.

To be fair, philosophy of mind HAS expanded its scope in recent years, but a large part of its history still involves what some have called the "<u>spectator theory of knowledge</u>" — an idea with a history stretching well beyond Alan Turing, or even John Locke, all the way to Aristotle. Most of those working in AI continue to operate under this outdated, limited model of human understanding as just an analog series of inputs and outputs, ordered by some imagined internal structure. If we've learned anything from the Inversion of our modern internet, it's that linear machine learning (no matter how much it's sped up) will always lack three key features of human understanding: *viz.* **contextualization, knowing-which, and creativity**. Sadly, once the novelty of weak A.I. systems like Siri and Alexa has worn off, I believe interest and funding in A.I. will continue to wane unless the entire project is reconceptualized. Thus the title of this talk - "The Looming AI Winter." I'd like to suggest philosophers would do well to set aside debates about Turing machines, Chinese Rooms, or "Harry" Lycan-thropes and turn instead toward theories which accentuate the non-linear, entangled aspects of human understanding - like those championed by post-structuralism, phenomenology, and pragmatism.

Even as the internet was still in its ARPANET infancy, Jacques Lacan was delving into the differences between "the Real" and the imaginary, Michel Foucault was reminding us about "The Treachery of Images," and Jean Baudrillard was laying out the stages of reality-destabilization that occur under a simulation — an analysis that DIRECTLY pertinent to phenomena like Read's Inversion and Lil' Miquela.

In *Simulacra and Simulation* (1981) Baudrillard explained how reality can be destabilized and replaced by a simulation. First, a simulation is seen as a faithful representation of reality — like a map or a portrait. Next, comes the perversion of that original reality as the symbol takes on a life of its own, through mass production. Caricatures are a one example of this — since they still "suggest" the original, but mostly consist in a mashup of social stereotypes and visual tropes. Third comes the stage in which we begin to "lose sight" of the original, as the simulation/caricature takes precedence over reality. (Growing up near one of the Disney theme parks gave me lots of experience with this stage.) Wherever a real original takes a backseat to a simulation — and begins to follow after it — this third stage has been reached. The completion of this process finally occurs once the simulation claims for itself its own reality. Even the faint echoes of the original have vanished, because the simulation has successfully supplanted it and has become a hyperreal *simulacrum*. This is perhaps the BROADEST sort of Inversion, one that can't even be expressed in quips like "up is down" because such categories have been obliterated.

Although approaches like Baudrillard's may be helpful for diagnosis, he fell a bit short when it comes to remedies… especially with regard to the consumption of mass media in our age of "fake news" and extreme political division. In his words,

> …the masses are also made of this useless hyperinformation which claims to enlighten them, when all it does is clutter up the space of the representable and annul itself in a silent equivalence. *And we cannot do much against this obscene circularity of the masses and of information. The two phenomena fit one another: the masses have no opinion and information does not inform them*. (Baudrillard 1985, 210)

Fortunately, there are other philosophers toward which the responsible **Netizen** can turn—like the 20th century American thinker John Dewey. Unlike Baudrillard, Dewey saw the sorts of ambiguities one finds online as a kind of **opportunity**. In his words, "Thinking begins in what may fairly enough be called a forked-road situation, a situation which is ambiguous, which presents a dilemma, which proposes alternatives." (John Dewey 1910, 11)

Rather than running from the uncertainty and/or different worldviews we encounter online, Dewey would've encouraged us embrace it all. By his lights, the two essentials of thinking were healthy skepticism and a willingness to inquire into what we don't know. He knew inquiry wasn't often easy, but he believed it was ALWAYS worthwhile.

> The easiest way [to cope with uncertainty] is to accept any suggestion that seems plausible and thereby bring to an end the condition of mental uneasiness. Reflective thinking is always more or less troublesome because it involves overcoming the inertia that inclines one to accept suggestions at their face value; it involves willingness to endure a condition of mental unrest and disturbance.

Reflective thinking, in short, means judgment suspended during further inquiry; and suspense is likely to be somewhat painful. (Ibid. 13)

That's why he advocated an education that promoted individual critical thinking. He understood well the kind of uncritical, mob mentality that human beings fall into whenever they flee from discomfort and uncertainty. "We talk about thinking for one's self. After all, the words 'for one's self' are superfluous or redundant. **It is not thought unless it is for one's self**." — *"Individuality in Education," 1922*

Dewey understood that successful democracies require citizens to have rich repertoires of cultural experience, critical apparatuses finely tuned to subtle political nuances, and the intellectual maturity to not be threatened by alternative points of view — while at the same time requiring a process sufficiently streamlined to address public needs with timeliness. And that's why he extolled the virtues of democracy — NOT merely as a political system — but as A WAY OF LIFE. In his words,

> …the task of democracy is forever that of creation of a freer and more humane experience in which all share and to which all contribute… to treat those who disagree — even profoundly — with us, as those from whom we may learn, and in so far, as friends… To cooperate by giving differences a chance to show themselves because of the belief that the expression of **difference is not only a right of the other persons but is a means of enriching one's own life-experience**, is inherent in the democratic personal way of life… **It is to realize that democracy is a reality only as it is indeed a commonplace of living**."— *"Creative Democracy: The Task Before Us," 1939*

Although we often hear appeals to celebrate what we share in common, Dewey also understood that *difference* is what drives a democracy forward, by forcing us to *reconstruct* those habits of thought to which each of us stubbornly clings into something we can more productively embrace *together*.

That may be easier said than done… but its the kind of task we MUST take up if we're going to keep our world from being "Inverted."

   The era of big data initiated by Web 3.0 has brought to light just how theory-laden information really is. Just as Wittgenstein's duck-rabbit example illustrates, seeing is always seeing-as. This idea is known as theoretical holism. As Michael P. Lynch has put it:

> there is no direct observation of the world that isn't at least somewhat affected by prior observations, experiences and the beliefs we've formed as a result... Like it or not, we can't do data analytics without theory. It's what gives us the context in which to pose questions and interpret the correlations we discover....the process of theorizing employs a composite of cognitive capacities, ones that when employed together make up understanding... understanding isn't piecemeal; it involves seeing the whole. (Lynch 2016, 159)

Recognizing context is a feature of human understanding still out of reach for AI. Programmers can provide a theory of interpretation, they've even had some success in having AI teach itself successful interpretive strategies, but these approaches inevitably bog down in the face of novel or ambiguous inputs. A good example of this can be found in the world of gaming, where Elon Musk's neural net, "OpenAI," has competed against teams of human experts in the Multiplayer Online Battle Arena game known as DOTA2.

In the last contest, held in April of 2019, the AI team defeated the human players in two of the three rounds. This was hailed as a great tipping point in AI. But, the devil's in the details. First, the rules of the game were heavily restricted, as only 18 of the possible 115 playable characters were available. But, more importantly, the humans gained a decisive victory in the third match by doing something completely unexpected—behaving erratically. Instead of playing like the savvy experts they were, the human players broke the norms of gameplay by pulling stunts only the greenest of "noobs" might attempt. The AI broke. Its characters began turning in circles as if they were locked in an endless loop of indecision. The human players were able to walk up and pick them off without resistance. The AI was simply incapable of adapting.

This example highlights some important points about context. It's fluid and dynamic. It shifts. Sometimes it even converges on other contexts in a single situation. Part of the maturation process of human intelligence is learning to attend to such subtleties. As John Dewey — in one of his grumpier sounding passages — wrote about drawing conclusions in an ever-changing world:

> Unless one is an idiot, one simply cannot help having all things and events suggest other things not actually present, nor can one help a tendency to believe in the latter on the basis of the former. The very inevitableness of the jump, the leap, to something unknown, only emphasizes **the necessity of attention to the conditions under which it occurs** so that the danger of a false step may be lessened and the probability of a right landing increased. — How We Think, *p. 26*

Machine learning, however complex, continues to operate on the same basic model as Searle's Chinese room, input-interpretation-output, and thus fails miserably at adaptation. This leads to the second feature of human understanding still out of reach for AI, *viz*. discernment or "knowing-which."

The human team of players was seamlessly able to bend, or even break, some rules on the fly. As a species we excel at breaking things. But, this is precisely what makes human intelligence so adaptable. Michael P. Lynch sums it up nicely:

> The person who truly understands, in the philosophical sense, is discerning not only the actual situation, but also why various hypotheses and explanations won't work as well as how to ask what would. (Lynch 2016, 171)

Being able to ignore correlations in data, even ones that appear quite strong, is one practical benefit of this capacity.

Consider the very strong — yet completely meaningless— correlation between Nicolas Cage films and the number of deaths due to falling into swimming pools between 1999 and 2009. I suspect few among us would conclude Nick Cage films have been causing drownings! To do so would be to commit the fallacy of confusing correlation with causation.

However, there are NUMEROUS instances of convolutional neural networks, like OpenAI, making this very mistake - which programmers call "optimizing for the wrong utility function." The classic example, which may or may not have actually happened, tells of a military neural network being taught to differentiate between enemy and ally tanks across a broad data-set of images. But since the enemy tanks were photographed on a cloudy day and the friendlies on a sunny one, the AI ended up identifying the tanks only when the weather conditions were perfect! A similar, documented account involved a NN learning to distinguish wolves from dogs. What it learned instead was that wolves stand on snow and dogs stand on grass! Another, more appalling, instance in 2015 involved Google's NN photo app mislabeling images of African Americans as "gorillas." And, in 2017, a review of research claiming a connection between iris texture and gender identification was found to have involved NNs optimizing for the utility function of mascara-no mascara, rather than discerning differences among irises.

What these examples demonstrate is that good old-fashioned AI, even at its pinnacle state of the art, falters in the face of ambiguous data points. But, such forked-road situations, as John Dewey asserted, are precisely where thinking begins. So why have AI enthusiasts remained so optimistic despite all these let-downs? The late phenomenologist Hubert Dreyfus offered an answer by highlighting four assumptions implicitly made by many AI optimists.

The first is the biological assumption that, on some level, the (human) brain functions like a digital computer, with a series of perceptual inputs, interpretive machine states, and behavioral outputs. Dreyfus leaned on phenomenological thinkers like Martin Heidegger and Maurice Merleau-Ponty to make this point, but could just have easily referenced Dewey, who in 1896 challenged the "reflex-arc" concept in the psychology of his day because it placed the parts of an act prior to the whole and failed to recognize that stimulus, movement, and response only make sense as an interpretation of an event after it has occurred. The second assumption Dreyfus noted was just a psychological version of the first, holding that "the mind, rather than the brain, functions like a digital computer, even if the brain doesn't happen to do so." (cf. Johan Nystrom-Persson) The third, epistemological, assumption is similar to the functionalism of Turing insofar as it sees the formalism of the brain/mind-as-computer analogy as still metaphorically explanatory in understanding intelligent behavior, even if not literally true. Meanwhile, the fourth and most crucial assumption, which Dreyfus considered ontological, is reminiscent of the picture theory of language associated with Bertrand Russell, G.E. Moore, and the young Wittgenstein insofar as information about the world is thought to be reducible to discrete facts.

By studying human expertise across a wide range of fields and comparing it to the development of AI expertise (including IBM's chess master Deep Blue and Watson of *Jeopardy!* fame) Dreyfus found that AI learning, even that of modern "deep" learning, operates more like humans advancing from novicey to competence than it does human expertise. Even though the development of machine learning was in part a response to the sorts of critiques Dreyfus made, neural networks still rely on rigid rules, have limited situational perception, and depend heavily on formulated routines. This helps explain why AI excels, and even outperforms humans, in very structured

settings like chess or standard DOTA2 matches. It may have greater processing speed, but it can't cope with deviations or consider new possibilities the way humans at the highest levels of skill can.

What those four assumptions point to is the hardest thing for many AI optimists to grasp. Simply put, the brain doesn't actually model the world. Since the brain is "in" the world, it has to sense it in real time to interact with it. As Loius Savain puts it, "There aren't enough neurons in the brain to store a model of the world." Moreover, because of its position in the the world, there is no Archimedean point from which it can view both the world and its model. Savain's metaphor for this mistaken view is an artist looking at both her subject and the canvas. To more accurately extend the metaphor, the brain is both the canvas and the artist, but as a part of the world it's also continuous with the subject matter.

This might be time for an important sidebar... really there are two problems plaguing AI. The first, more primordial one is that which I've been discussing so far - *viz.* that AI optimists are confused about how humans come to understand. The second, which also permeates pop culture, is that if, or when, AI achieves sentience, it's going to look anything like human sentience. This is highly unlikely. Just as human beings were able to engineer flight but achieved it in a way radically different from the source material, *i.e.* birds, any realized strong AI will be equally different from its human models. The silver lining here is that the sci-fi trope of AI machines suddenly becoming self-aware revolutionaries is simply not in the cards. Westworld's Dolores and her awakened host compatriots are acting from recognizably <u>human</u> motivations like vengeance, maternal instinct, and empathy. To pilfer a phrase from Nietzsche, these replicants are "human, all too human."

This point is important because it highlights how AI theorists and developers have locked themselves in a conceptual prison of their own design. The way out, and thus forward, may turn out to be a re-imagining of human understanding. Let's try an exercise to illustrate the point. I'm going to utter two sentences and I want you to close your eyes and picture the scenes. Are you ready? Here's the first sentence: "They gave us a hearty welcome." Can you picture the scene this sentence calls to mind? Now try the second sentence: "The gave us a cordial reception." Did the scene change? If you're like most people, the first sentence probably felt more intimate while the second sentence likely elicited images of a more formal affair. The reason for this difference has to do with the conceptual history of these words since the Norman invasion of 1066. Without getting too technical, the first sentence derives from the Anglo-Saxon of the occupied, peasant class, while the second derives from the French of the Norman ruling class. Though none of us have a connection to this event, the conceptual distinction lives on in the language.

There's a similar issue in the history of philosophy regarding human experience and understanding. When discussing the acquisition of knowledge, vision continues to be the dominant interpretive paradigm. The words we use - like reception, perception, conception, or inception - almost all derive from Romance-language variants of the Latin for 'taking in.' The history of empiricism, from Aristotle to Locke to contemporary philosophers has, by and large, been held captive by these conceptual barriers. Dewey was heavily critical of this tradition, especially in his later works and dubbed it "the spectator theory of knowledge."

For both Dewey and Dreyfus, experience referred to something much more engaged, embodied, and situated within the world than the spectator theory concepts would lead one to

believe. Each man's project could be viewed as an attempt to reunite the pieces of experience that had been, in Dewey's phrase, "torn asunder." For Dewey, this meant that:

> experience is [just] heightened vitality. Instead of signifying being shut up within one's own private feelings and sensations, it signifies active and alert commerce with the world. (LW 10:25)

As such, it's shot through with inference and reflection:

> taken free of the restrictions imposed by other concepts, [experience] is full of inference. There is, apparently, no conscious experience without inference; reflection is native and constant. (MW 10:6)

Dreyfus agreed with Dewey that there's no sharp distinction to be found between action and thought, though he came to the conclusion from another direction. He was critical of philosophers, as well as those working in AI, for treating "all understanding as an epistemological problem, as a question of theoretical knowledge." He argued there was a more primordial, practical understanding that was irreducible to proposition. In his words,

> Although practical understanding—everyday coping with things and people—involves explicit beliefs and hypotheses, these can only be meaningful in specific contexts and against a background of shared practices...What makes up the background is not beliefs, either explicit or implicit, but habits and customs, embodied in the sort of subtle skills which we exhibit in our everyday interaction with things and people. (Dreyfus 2014, 130)

In the view shared by Dewey and Dreyfus, lived experience is simultaneously a cognitive and practical affair, with conception, perception, deliberation, and action occurring in chorus, NOT sequentially. Dewey called this inquiry, which for him was a wholly organic process of enriching one's contexts by transforming and freeing up the world in which experience was always situated. Dreyfus employed Heidegger's *verstehen*, or vital understanding, to make the same point about "the non-cognitive precondition of all understanding" (Ibid. 133).

Michael P. Lynch underscores this Dewey-Dreyfus connection in highlighting that third feature of human understanding that AI has yet to achieve, namely, creativity. As he puts it:

> In order to understand, one must first come to understand, and it is this coming to understand that is a [creative] act... Creative ideas are valuable to the person's cognitive workspace. They move things forward on the conceptual field on which they are currently playing... One acts by opening the door, and then one is acted upon by seeing what lies beyond. Understanding is a form of disclosure. (Lynch 2016, 177)

In recent years there have been claims regarding the supposed creativity of AI. In 2015, Google's DeepDream neural network made headlines by producing haunting images built from its database of millions of internet images. More recently, a pair of Stanford students were celebrated for teaching a neural network to generate memes - which they called "dank learning." You can take a

look at some of the memes to the right and decide for yourself if they'd be funny without the knowledge that they were created by AI. You might even find that knowledge enhances the humor. But would they qualify as creative acts of "coming to understand" as outlined by Dewey, Dreyfus, and Lynch?

The financial incentives to hyperbolize this sort of thing are obvious. What's troubling is not that such claims about AI creativity may be fake, but rather the wrong-headed view of understanding they perpetuate. In Tristan Greene's write up on "dank learning" for TheNextWeb.com we get a prime example. He writes,

> Just like any other neural network, it bounces an idea around between different learning layers until it comes up with something similar to what it's been trained on.

As he puts it:

> People learn the same way. (Greene 2018)

Except… they don't!

Sources

Alexander, Thomas M. *John Dewey's theory of art, experience, and nature: The horizons of feeling*. SUNY press, 2012.

Dewey, John. *How We Think*. Heath & Co. Publishers, 1910. [MW 10]

Dewey, John. *Art as Experience*. Minton, Balch & Co. 1934. [LW 10]

Dreyfus, Hubert L. *Skillful coping: Essays on the Phenomenology of Everyday Perception and Action*. OUP Oxford, 2014.

Dreyfus, Hubert L., and Stuart E. Dreyfus. "What Artificial Experts Can and Cannot Do." *AI & Society* 6.1 (1992): 18-26.

Greene, Tristan. "Stanford Researchers Taught AI to Make Dank Memes," *thenextweb.com*, June 13, (2018).

Gwern, Branwen. "The Neural Net Tank Urban Legend." (2011) https://www.gwern.net/Tanks.

Hart, Matthew. "Westworld's Version of the A.I. Singularity Will Never Happen in Real Life." *Nerdist.com*, December 6, (2016).

Lynch, Michael P. *The Internet of Us: Knowing More and Understanding Less in the Age of Big Data*. WW Norton & Company, 2016.

Kuehlkamp, Andrey, Benedict Becker, and Kevin Bowyer. "Gender-from-iris or gender-from-mascara?" 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017.

Nystrom-Persson, Johan. "Dreyfus and Bostrom. Four AI assumptions and two books." *Monomorphic.org*, April 23 (2017).

Petrarca, Emilia. "Body Con Job Miquela Sousa has over 1 million followers on Instagram and was recently hacked by a Trump troll. But she isn't real." *The Cut*, May 14, 2018.

Read, Max. "How much of the internet is fake? turns out, a lot of it, actually." *New York Magazine*, December 26. (2018).

Savain, Loius. ""The World Is its Own Model or Why Hubert Dreyfus Is Still Right About AI" *Medium.com*, October 5 (2017).

Simonite, Tom. "When it comes to gorillas, google photos remains blind." *Wired*, January 13 (2018).