OSCAR SCHWARTZ | WED 25 JUL 2018 11.00 BST LAST MODIFIED ON WED 25 JUL 2018 21.39 BST

'The discourse is unhinged': how the media gets AI alarmingly wrong

Social media has allowed self-proclaimed 'AI influencers' who do nothing more than paraphrase Elon Musk to cash in on this hype with low-quality pieces. The result is dangerous

In June of last year, five researchers at Facebook's Artificial Intelligence Research unit published an article showing how bots can simulate negotiation-like conversations.

While for the most part the bots were able to maintain coherent dialogue, the researchers found that the software agents would occasionally generate strange sentences like: "Balls have zero to me to me to me to me to me to me to me to."

On seeing these results, the team realized that they had failed to include a constraint that limited the bots to generating sentences within the parameters of spoken English, meaning that they developed a type of machine-English patois to communicate between themselves. These findings were considered to be fairly interesting by other experts in the field, but not totally surprising or groundbreaking.

A month after this initial research was released, Fast Company published an article entitled AI Is Inventing Language Humans Can't Understand. Should We Stop It?. The story focused almost entirely on how the bots occasionally diverged from standard English – which was not the main finding of the paper – and reported that after the researchers "realized their bots were chattering in a new language" they decided to pull the plug on the whole experiment, as if the bots were in some way out of control.

Fast Company's story went viral and spread across the internet, prompting a slew of content-hungry publications to further promote this new Frankenstein-esque narrative: "Facebook engineers panic, pull plug on AI after bots develop their own language," one website reported. Not to be outdone, the Sun proposed that the incident "closely resembled the plot of The Terminator in which a robot becomes self-aware and starts waging a war on humans".

Zachary Lipton, an assistant professor at the machine learning department at Carnegie Mellon University, watched with frustration as th

According to Lipton, in recent years broader interest in topics like "machine learning" and "deep learning" has led to a deluge of this type of opportunistic journalism, which misrepresents research for the purpose of generating retweets and clicks – he calls it the "AI misinformation epidemic". A growing number of researchers working in the field share Lipton's frustration, and

worry that the inaccurate and speculative stories about AI, like the Facebook story, will create unrealistic expectations for the field, which could ultimately threaten future progress and the responsible application of new technologies.is story transformed from "interesting-ish research" to "sensationalized crap".

Exaggerated claims in the press about the intelligence of computers is not unique to our time, and in fact goes back to the very origins of computing itself.

In February 1946, when the school bus-sized, cumbersome Electronic Numerical Integrator and Computer (Eniac) was presented to the media at a press conference, journalists described it as an "electronic brain", a "mathematical Frankenstein", a "predictor and controller of weather" and a "wizard". In an attempt to tamp down some of the hype around the new machine, renowned British physicist DR Hartree published an article in Nature describing how the Eniac worked in a straightforward and unsensational way.

Much to his dismay, the London Times published a story that drew heavily on his research titled An Electronic Brain: Solving Abstruse Problems; Valves with a Memory. Hartree immediately responded with a letter to the editor, saying that the term "electronic brain" was misleading and that the machine was "no substitute for human thought", but the damage was done – the Eniac was forever known by the press as the "brain machine".

It was a similar story in the United States after Frank Rosenblatt, an engineer at Cornell Aeronautical Laboratory, presented a rudimentary machine-learning algorithm called the "perceptron" to the press in 1958. While the "perceptron" could only be trained to recognize a limited range of patterns, the New York Times published an article claiming that the algorithm was an "electronic brain" that could "teach itself", and would one day soon "be able to walk, talk, see, write, reproduce itself and be conscious of its own existence".

While the giddy hype around AI helped generate funding for researchers at universities and in the military, by the end of the 1960s it was becoming increasingly obvious to many AI pioneers that they had grossly underestimated the difficulty of simulating the human brain in machines. In 1969, Marvin Minsky, who had pronounced only eight years earlier that machines would surpass humans in general intelligence in his lifetime, co-authored a book with Seymour Papert proving that Rosenblatt's perceptron could not do as much the experts had once promised and was nowhere near as intelligent as the media had let on.

Minsky and Papert's book suffused the research community with a contagious doubt that spread to other fields, leading the way for an outpouring AI myth debunking. In 1972, the philosopher Hubert Dreyfus published an influential screed against thinking machines called What Computers Can't Do, and a year later the British mathematician James Lighthill produced a report on the state of machine intelligence, which concluded that "in no part of the field have the discoveries made so far produced the major impact that was then promised".

This trough of disillusionment ushered in what has since been called the first AI winter, a period in which funding for research in the field dropped off almost entirely. The media, which had drummed up so many inflated expectations for "electronic brains", also lost interest. While there were small resurgences in the 1980s and 1990s, AI was more or less a topic relegated to the realm of corny sci-fi novelists – computer scientists often avoided the term artificial intelligence altogether for fear of being viewed as "wild-eyed dreamers".

The ice of AI's first winter only fully retreated at the beginning of this decade after a new generation of researchers started publishing papers about successful applications of a technique called "deep learning".

While this was fundamentally a decades-old statistical technique similar to Rosenblatt's perceptron, increases in computational power and availability of huge data sets meant that deep learning was becoming practical for tasks such as speech recognition, image recognition and language translation. As reports of deep learning's "unreasonable effectiveness" circulated among researchers, enrollments at universities in machine-learning classes surged, corporations started to invest billions of dollars to find talent familiar with the newest techniques, and countless startups attempting to apply AI to transport or medicine or finance were founded.

As this resurgence got under way, AI hype in the media resumed after a long hiatus. In 2013, John Markoff wrote a feature in the New York Times about deep learning and neural networks with the headline Brainlike Computers, Learning From Experience. Not only did the title recall the media hype of 60 years earlier, so did some of the article's assertions about what was being made possible by the new technology. "In coming years," Markoff wrote, "the approach will make possible a new generation of artificial intelligence systems that will perform some function that humans do with ease: see, speak, listen, navigate, manipulate and control."

Since then, far more melodramatic and overblown articles about "AI apocalypse", "artificial brains", "artificial superintelligence" and "creepy Facebook bot AIs" have filled the news feed daily.

Lipton, a jazz saxophonist who decided to undertake a PhD in machine learning to challenge himself intellectually, says that as these hyped-up stories proliferate, so too does frustration among researchers with how their work is being reported on by journalists and writers who have a shallow understanding of the technology.

What Lipton finds most troubling, though, is not technical illiteracy among journalists, but how social media has allowed self-proclaimed "AI influencers" who do nothing more than paraphrase Elon Musk on their Medium blogs to cash in on this hype with low-quality, TED-style puff pieces. "Making real progress in AI requires a public discourse that is sober and informed," Lipton says. "Right now, the discourse is so completely unhinged it's impossible to tell what's important and what's not."

Lipton is not the first person to express concern about the new AI hype cycle and where it is taking the field. Last year, the pioneering roboticist Rodney Brooks wrote an article criticizing the "hysteria about the future of artificial intelligence" for MIT Technology Review. In 2013, New York University professor Gary Marcus wrote an article for the New Yorker in which he argued that the hype will create unrealistic expectations followed by disillusionment leading to another AI winter.

But for Lipton, the problem with the current hysteria is not so much the risk of another winter, but more how it promotes stories that distract from pressing issues in the field. "People are afraid about the wrong things," he says. "There are policymakers earnestly having meetings to discuss the rights of robots when they should be talking about discrimination in algorithmic decision making. But this issue is terrestrial and sober, so not many people take an interest."

In March last year, Lipton started his own blog to try to counter-balance and deconstruct some of the most damaging sensationalist AI news. So far, he has called out a "low-quality" profile of Elon Musk in Vanity Fair and an uncritical overview of Anthony Levandowski's so-called AI church, among others.

The blog has received some attention from the media and has a regular readership, but Lipton knows that his influence is limited. "What really needs to happen is better training of journalists and more integrity," he says. "Until that happens, my blog is just a pebble in the rapids of crap. I'm not altering the direction of the stream."

Joanne McNeil, a writer who examines emerging technologies, agrees that there is a problem with uncritical, uninquiring tech journalism, and often uses Twitter to make fun of Terminator-style articles. But at the same time, she is weary of pointing the finger solely at journalists and believes that one of the causes of AI hype is an uneven distribution of resources.

"If you compare a journalist's income to an AI researcher's income," she says, "it becomes pretty clear pretty quickly why it is impossible for journalists to produce the type of carefully thought through writing that researchers want done about their work." She adds that while many researchers stand to benefit from hype, as a writer who wants to critically examine these technologies, she only suffers from it. "There are few outlets interested in publishing nuanced pieces and few editors who have the expertise to edit them," she says. "If AI researchers really care about being covered thoughtfully and critically, they should come together and fund a publication where writers can be suitably paid for the time that it takes to really dig in."

While closer interaction between journalists and researchers would be a step in the right direction, Genevieve Bell, a professor of engineering and computer science at the Australian National University, says that stamping out hype in AI journalism is not possible. Bell explains that this is because articles about electronic brains or pernicious Facebook bots are less about technology and more about our cultural hopes and anxieties.

"We've told stories about inanimate things coming to life for thousands of years, and these narra-

tives influence how we interpret what is going on now," Bell says. "Experts can be really quick to dismiss how their research makes people feel, but these utopian hopes and dystopian fears have to be part of the conversations. Hype is ultimately a cultural expression that has its own important place in the discourse."

Ultimately, Lipton agrees that there is a place for speculative writing about AI and he acknowledges how imagination and emotion can motivate inquiry in the field. "But I also think that boundary between wild speculation and real research is a little too flimsy right now," he says. "As history shows us, this is a boundary that needs to be monitored so we can distinguish between what's important in the here and now and what's just fantasy."