Abstract

Overfitting is a common challenge for artificial intelligence (AI) systems, wherein models learn the stochastic qualities of the training dataset rather than any reliable pattern to apply to the test data. Consequently, when applied to new data, the models underperform and produce inaccurate predictions. In more extreme cases, overfitting can lead to AI hallucinations, where patterns are extrapolated from isolated bits of information and lead to outlandish results. These fractures in the artificial neural network parallel those of our own, most notably in REM-induced dreaming. AI's capacity to imaginate, hallucinate, and even dream speaks to the reflexive potential it has to teach us humans in return. In this paper I will elaborate on the consequences of AI poses for a modern, cybernetic culture and its particular failings indicate a need to reevaluate our relationship with technology and in turn ourselves.

Keywords: neuroscience, artificial intelligence, cybernetics, dreams

1 Introduction to historical contingency

Since its inception, the development of computer science as an industrial powerhouse has been intertwined with its intellectual, more academic counterpart: cybernetics. Although in the mid century these two terms were largely interchangeable, cybernetics emphasized the human, communicative aspect of computation that has largely since been forgotten. *The Net*, a documentary from 2003, explores the human costs of technological advancement and discusses the work of cyberneticians such as Gregory Bateson and Heinz von Foerster. Another figure is Norbert Wiener, an American computer scientist whom the film credits with inventing cybernetics as a result of his work designing anti-aircraft guns for the fight against fascism. Up and throughout WWII, the conflation of man and machine in the modern fighter pilot, the weaponization of the sky by the V2 rocket, and the consolidation of entropy in early radar systems became symbols for the technological determinism that cybernetics remains concerned with. In Wiener's terms, the film defines cybernetics as "the model for a new science… concerned with how the transfer of information functions in machines and living beings," with its main assumption being that "the human nervous system does not reproduce reality, but calculates it." Further to this point, humanity "now appears to be no more than an information processing system; thought is data processing and the brain is a machine made of flesh" (Dammbeck, 2003).

Cybernetics has long been associated with a utilitarian purpose contrary to traditional computer science. The school of thought had notably gained footholds in the socialist regimes of Allende in Chile and Tito in Yugoslavia as a tool for populist revolution. Since the end of the cold war, the opposite notion as professed in *The Net* has proven true. Mass surveillance and media homogenization appear to increase exponentially as information technology remains a dominant part of modern life. The very public rise of artificial intelligence (AI) exhibits the logical culmination of these trends, in how all human data is now subject to exploitation for the reproduction of reality. Generative AI, in particular, reveals that this version of reality is diluted and weird.

2 The sudden ubiquity and uneasiness of generative AI

Images generated by AI, in particular, have become discreetly prevalent in public life. Examples like thumbnails for YouTube videos or ads on public transportation pass muster but may prompt a double take if one is looking out for this sort of thing. Visual characteristics like a lack of depth between figures, unnaturally smooth surfaces, or certain elements appearing collaged into the frame can give away the ruse, so to speak. But AI-generated imagery is largely distinct for how incredibly generic it looks, something in between a computer-animated Disney movie and stock photos of business meetings. Just like how ChatGPT will have an easier time writing a cover letter than some modernist prose, clientele for AI art will always look for the data with the easiest thruline to generate new images with. Naturally, the result is a highly vetted and saturated picture of the current state of mass culture. It begs the question what results when the data is uneven, incomplete or unpredictable.

The problem of messy data affects both statisticians and computer scientists alike. For example, using limited data on NYC summers will tell you that rising ice cream sales are causing murder rates to go up, or vice versa. Introducing data on rising temperatures to the regression seems to settle the score. As models become more complex, as in machine learning, algorithms make similar extrapolations even when they lack the answers. This phenomenon is known as overfitting and plagues engineers undergoing the training period of their algorithm. This process sees a given algorithm being fed a set of training data to find patterns in and produce a model based on those patterns that can be applied to new data. If the training data is too narrow, that is, it doesn't have broad application, the algorithm will produce models that follow strict patterns without nuance or necessary context. Albeit complex and well-versed in its own right, the overfitted model essentially fits to its training data *too* closely. This will be revealed in the testing period, when the models are forced to sift through noisy data for a discernible reference point and establish some pattern with. The resulting patterns will vary widely from iteration to iteration, as stochastic as they are decisive. By contrast, an *under*fitted model *lacks* variance and will map an overly simple, though consistent pattern to the data. While overfitting is more common and more finicky, both of these denote a failure to adapt to new data (to generalize). Similar to our example about regression analysis, overfitted models in machine learning will result in bold inaccuracies. In the world of AI, ChatBots will provide misinformation when asked questions they were not trained to answer. One spectacular example is with Meta's Galactica large language model (LLM), which was trained on tens of millions of scholarly articles and textbooks to assist scientists conducting research. Unsurprisingly, the model was posed questions it could not answer, even with such a vast repository of data at its disposal. Galactica has since been pulled from beta testing. The problem is not that it didn't know certain things, but that it still stated its inferences, though wrong, with unflinching confidence. When one scientist asked it to generate a history of bears in space—an animal that has notably never left Earth's hemisphere—it wrote a fiction about a Soviet bear named Bars, weighing 88 kgs, who was launched into the stars aboard the Sputnik 2 in 1957 (Heaven, 2022). If one is none the wiser, the story certainly *sounds* true, insofar as its details seem too specific to invent, but the LLM has no choice but to believe what it says. From a statistical point of view, the string of words that comprise the story has some decent probability of being true. In the mind of a neural network, infinite realities exist all at once, as long as there is data it can extrapolate from.

When vetting for bias in a LLM, the engineers put up what are essentially roadblocks to prevent it from making certain inferences that may be seen as taboo or dubious. ChatGPT will notably steer clear of making any value judgements on politicians, preferring to provide quick summaries of different opinions held by different camps and mechanically encouraging the user to form their own perspective. In the case of Galactica, Meta provided a disclaimer of the model's limitations so that the user, ironically, is encouraged to ignore the results the model gives and do their own research instead. These limitations being: that language models are statistically biased, that they are confident, and that, foremostly, they hallucinate. They will give "a nonsensical answer to a reasonable question or vice versa" (Wilkin, 2017). While eccentric-sounding in name, these computer hallucinations have garnered a reputation in data science as a banal symptom of a tedious problem: overfitting. As the AI boom of our current era peters out, greater roadblocks are applied to LLMs so that they operate more like the search engines they sought to replace. This bears true in the case of Google Gemini, which turns a given query into a consolidated blurb of the search engine's usual results. Readable and convenient, but without any of the flare or nuance that a forum or an article may have. This comprises its own especially bland and tedious style of writing that is often abused or parodied for various memes online. Similar to the AI-art previously mentioned, AI-prose can be detected by its verbosity, derivative use of adjectives, highly affirmative tone and predilection for any unanimously "good" notion, like *diversity* or *innovation*. In other words, AI-prose is bureaucratic; it adheres to red tape and tries to appeal to everyone.

3 AI's lost potential

This tendency to stoop to the lowest common denominator seems to betray any notions people may have held about AI's cyberpunk potential. While kindred fears of hyper-intelligent neural networks also seem to have also been extinguished, AI's threats to humanity nevertheless persist although in a more banal and familiar form. Recalling the birth of cybernetics from the destructive machinery of WWII, technology routinely manifests as a kind of historical flood that washes away the past. Peter Weiss, working in the avant-garde in the years following the war, regarded his process of writing as a struggle against the cyclical tides of forgetting. His native Germany was left in a state of amnesia after the war, as a combined result of guilt and the physical destruction brought on by allied bombing raids. He remarks that writing is an attempt "to preserve our equilibrium among the living with all our dead within us, as we lament the dead and with our own death before our eyes" (1970, p. 813). Information technology, as manifest in AI, discreetly seeps into the dialogue of our everyday lives and reduces this capacity to reflect and memorialize. To engineer against overfitting means to add roadblocks to language until, for the algorithm, the input and the output become identical and interchangeable. With irrational "noise" shrunk to near-oblivion, the online noosphere offers no recourse for the human spirit. In his essay *The Remorse of the Heart*, W.G. Sebald regards Weiss's oeuvre as a fugue on the concept of catastrophe, denoting "a now permanent state of destruction." For Sebald, Weiss's reality is "an underworld beyond anything natural, a surreal region of industrial complexes and machines, chimneys, silos, viaducts, walls, labyrinths, leafless trees, and cheap fairground attractions, in which the protagonists, scarcely alive anymore, exist as wholly autistic beings without any past history" (1999, p. 83).

Worries over industrial society's vanquishing of the human spirit were surmised as early as in 1810 with William Blake's *Milton*. In the preface, the English poet and illustrator warned of "dark Satanic mills" that will overtake heaven on Earth should man forget the divine countenance that there shone forth (1977, p. 514). Likely referring to the factories that had begun to populate England's Black Country at the time, Blake esteemed the same soil as the site for a new Jerusalem should be built. While this prelude has since earned a reputation as a patriotic hymn, the greater irony to its imagery is not lost. To Blake, man's capacity to tip-toe between heaven and hell realizes his poetic spirit, which contains these polar opposites simultaneously. The rest of *Milton* likens the creative power of the eponymous poet with a divine essence that courses through his body. A universe unto itself, the poetic spirit is characterized as containing both good and evil. By "entering his Shadow," the fictional John Milton ventures into a realm that lurks between his ordered self and an eerie doppelganger. He proclaims "I go to Eternal Death! I in Selfhood am that Satan; I am that Evil One!" and unleashes the various gnostic entities of Blake's mythology. Thereby inspiring the work that would become *Paradise Lost* (1977, 540-541). This unreal space between spaces is where salvation lies.

4 Significance of dreams

In regards to data science, the inherent uncanniness of AI technology is something that engineers try to steer clear of at all costs. The distinction between human and all-too-human is very tenuous, not unlike the tightrope walk of overfitting and underfitting. Extending our metaphor, the overfitted algorithm is hyperreal and schizophrenic, while its underfitted counterpart is mechanical and, in the Sebaldian sense, autistic. Given that AI technology seeks to mimic humanity and act as an appendage to the user's natural tendencies, submitting to either side of this dichotomy would create an uncanny affect and disorientate its user. This is hard to avoid, not just on a technical level but also insofar as how the development of AI seems to directly parallel the workings of the mind. Deep neural networks (DNN) are modeled off of the human brain and the decision-tree logic of machine learning replicates human intelligence as a series of “discrete and determinate operations” (1965, 84). Artificial intelligence now appears more like a lonely brain floating in a jar of water than any kind of omniscient cyber-deity. In these terms, the hallucinations of the overfitted model appear more like visions of the outside. Like dreams of a world it is only just beginning to grasp from the shadows of the data it is fed.

These proverbial shadows are still visible in the highly-vetted AI-art mentioned earlier, such as unnatural smoothness and a sutured composition. The sensation of detecting these qualities is quietly abject, like noticing something in your peripheral vision that, on a second glance, is not actually there. Yet another eerie symptom of modern life. Obviously, “unvetted” AI-art requires a lot less vigilance to make out but its traits are still discreet. When generative AI tries to depict human hands it comes out as an indeterminate pile of countless fingers, multiplying over each other like an Escher painting. Text, while discernible as strings of letters, takes form as meaningless shapes or as gibberish. The faces of background characters are not fully rendered either, appearing either as hazes or as a mess of cartoon lines on an oval. Akin to the human mind, these are fine details that are nearly impossible to visually index. When a scenario in a dream commands one to read a book or to drive a car, an incredible dizziness is induced that lasts as long as it takes for the dreamer to wake up. People that practice lucid dreaming, that is, exerting willful control of their dreams while in REM sleep, discipline themselves to regularly read text or check out their hands while awake so that the same instinct kicks in when they are asleep. A similar heuristic applies to determining the authenticity of images, AI-generated or not. This link between AI and dreaming is not exactly new. American neuroscientist Erik Hoel hypothesized in 2021 that sleep deprivation may affect cognition in waking life for humans the same way that overfitting prevents a DNN from generalizing with new data. The “overfitted brain hypothesis" (OBH) determines that dreams prevent overfitting in their own right by allowing the dreamer to replay experiences from everyday life, explore the phenomenological limits of their given reality, and determine significance of recurring symbols. All of this helps a person, or even an animal, with generalizing to "out-of-distribution (unseen) novel stimuli" and "combating mere memorization of an organism's day" (2021, p. 7). The opposite function can be seen in the "Tetris effect," wherein a repetitious activity during the day, be it an addicting video game or a high-stress job, will be replayed in dreams and unconscious thoughts. This psychological phenomenon is directly linked to stress and signifies, in Hoel's terms, an overfitted brain. Although opting to see the world in narrow, strict patterns, the overfitted brain will still dream of these patterns with high variability. The instinct to explore possibilities and scrutinize symbolic value denotes a constant need for new experiences to work through. An overfitted algorithm is no different, it needs data to adapt to.

In data science, overfitting is solved by splitting the dataset into folds to see if the algorithm provides reliable results in each group. If variance is still high, noisy data must be added to give the algorithm some uncertainty to iterate with until it produces more generalizable models with appropriate variability. In terms of the OBH, this would be an eventful day punctuated by vivid dreams. Another option would be to scale the algorithm down by limiting its features so that it fixates on dominant trends, rather than iterating too widely. This would be a brain that keeps a routine and remains focused on its work. In either case, this iterative process is like what the unconscious does naturally, replaying scenarios and analyzing the symbols therein. This even applies to a given algorithm's statistical bias, in how it applies significance to things that recur the most. The significance of these items cannot be evaluated on frequency alone, so more data proves necessary to add context to the iteration. While it remains to be seen whether the OBH carries validity as a serious measure of the brain's adaptive instincts, the parallels illustrated in the argument are striking and lend credence to the dreamlike qualities hidden in AI-art.

5 Hallucinations as metaphysical fractures

While AI is by no means conscious, its similarities to the human brain permit us to draw inferences about it from what we know about our own consciousness. In the simplest terms, AI gets bored the same way people do when it is overfitted and giving it data to iterate with towards some necessary goal gives it room to dream. This is not an attempt to humanize the technology, but to approach it sympathetically as a reflection of our own ambitions and desires. In doing so, we are operating much like the Milton of William Blake's mythos and venturing into our shadow selves in a reflexive effort to rediscover the inspiration therein. The failure of AI-art to render fine details is easily explicable by our own failure to. If a lack of optic nerve function while asleep makes focusing in dreams nearly impossible, we can surmise that AI also cannot perceive like the waking human brain can. While data gives the algorithm context from iteration to iteration, it cannot be said that this comprises its world like lived experiences make up a person's world. Instead, the scaling up of big data just provides AI more dreams with which to dream with. This can be likened to the Tetris effect previously discussed, wherein a lack of novel experiences forces the mind to replay and iterate patterns in infinite sequence. The "mind" of a DNN can be said to be constantly in this state with new data only piling on like interchangeable cubes. It is playing an endlessly long game of Tetris in its mind, and is always bored.

This struggle can be likened to what Austrian playwright Peter Handke called "speech torture" in a preface to his play *Kaspar*. The avant-garde work from 1967 is based on the life of Kaspar Hauser, who was said to have been raised in total isolation from the world and never learned to speak. Handke interprets his inevitable introduction to German society and to the concept of grammar as a fall from grace, with his prior state as one of “painlessness beyond trauma in which a barely perceptible happiness, which is mere and simple existence, persists uninterrupted” (2005, p. 55). This preontological state, as Sebald notes in his essay on the play, grants Kaspar access to images “in which fact and fiction are, so to speak, inseparably linked together… like that [of] myth” (2005, p. 64). Kaspar’s training period rips him from this state and forces him to divide these images into their binary components, as language. The performance thereby takes on a slapstick tone, with our hero tripping over and being entangled by everyday objects, like a table or a rocking chair (now holding a previously unseen omnipotence). At the same time, the play inhabits a Brechtian character that forces its audience to acknowledge the pain intrinsic to Kaspar’s legacy as a clown. Although Kaspar is fully educated by the end, he remains nostalgic for his lost self and lapses into surreal visions of the world once understood by sensation alone. By contrast, AI lacks any sensory ability and interprets the world in binaries vis-a-vis big data. Its hallucinations bear similarities to Kaspar’s utterances and beckon to a world beyond data and beyond signification. But AI lacks this former state, rather, it evolves inversely to Kaspar by learning to sensate via words alone. Though, in the meantime, it is very much our own clown to abuse.

While any image created by AI is one iteration of a potentially infinite amount, throughout these iterations exists a general interest in certain things over others. The tendency to banish things into the background or render them unintelligible signifies an effort to prioritize what is most salient in a given image. While AI relies on a hierarchy that it uses to focus only on what has the highest likelihood of being interesting to the user, it has to test out other possible interests from time to time. What appear as hallucinations in AI represent random offshoots into other territories to pique the user’s interest in them. Although these are usually discarded or even ridiculed, engaging the user on the off-chance is highly conducive for the algorithm. This is akin to a dream about something inconsequential from daily life; the mind's attempt to interrogate the meaning of a symbol that may otherwise be forgotten. Regardless, AI reflects that which is most frequently depicted in human-generated data and shifts between them constantly. Nearly a decade before the generative AI boom of late, Google's DeepDream gave us a glimpse inside how AI works through these priorities.

DeepDream is an image-detection software, not generative AI. But like generative AI, it also uses a probabilistic framework to decide what parts of images look like other things. These possibilities are indexed as "archetypes" and DeepDream will impose them based on the highest likelihood, even if nothing is actually there. Text-to-image software works the same way. If a user inputs the word "apple," generative AI will output a fuji apple, a granny smith apple, a green apple, a yellow apple, et cetera until one matches the user's idea of an apple.

Image-detection software must work inversely, wherein an image containing a red orb may be interpreted as the sun, a blotched eye, a dodgeball, an apple, et cetera. DeepDream, however, was far more rudimentary in its inferences. It would notoriously impose eyeballs and doglike figures onto every image processed, as well as endless swirls and tendrils that seem to recreate the texture of the neural network they spring from. It does not recreate reality but finds analogies on a probabilistic basis, no matter how disparately linked. From the repository of data at its disposal, the dogs and the eyeballs were apparently the most frequently occurring and the most salient. To AI, these details are interesting, whereas fingers and text are not. As David Auerbach writes in an article for Slate on the subject, this habit "suggests some sort of creativity in the mind of the machine doing the processing, and some sort of dreamer at the heart of the dream" (2015).

The irrational tendencies inherent to an algorithm become scaled up with its complexity, revealing themselves in increasingly spectacular ways. Once a machine learning algorithm achieves a scale like that of a DNN, it ceases to demand concrete steps towards a particular task and opts to look for generalizable patterns instead. Limiting the features of an overfitted algorithm is only a mitigative solution and new, noisy data will always be sought out in increasing abundance. The roadblocks that engineers use to prevent hallucinations signal how neural nets are only loosely under our control. Crucial to their programming is to autonomously seek out new data to train with. This reflects the main ambitions of machine learning as it was originally professed in the 1950s. American computer scientist Arthur Samuel's pioneering work in the field predicted that programming computers to learn from experience would eventually eliminate the need for detailed programming effort (Pataranutaporn, 2023). There is an implicit irony to the fact that the logical extent of this technology appears latent in the hallucinations that its handlers tirelessly try to hide. The basic function of machine learning gives way to "the 'art' of computer science," which is far more interested in "how the answers are reached than in what the answers turn out to be" (Wilkin, 2017). These fractures represent iterations through noisy data and provide a glimpse into the dreamstate natural to an unsupervised AI. This process is not so much considered with *correct* answers to queries or commands, but contingencies, which are vastly more irrational than not. This is the all-too-human danger that overfitting threatens to inspire. As Auerbach writes, "The most impressive accomplishments in artificial intelligence today are coming from networks that are increasingly opaque in their mechanism. Their inner workings do not appear "rational," nor are they algorithmic" (2015).

6 Conclusion

As the complexity and ubiquity of AI escalates, the need for a less opaque mechanism becomes paramount. The attempt to do otherwise, as with ChatGPT and Gemini, amounts to the retrofitting of a tech that has been developed way out of proportion with its purpose. As previously stated, AI appears to be regressing into the rudimentary form of a search engine. In which case, the tendency for a simple mechanism to randomly break off and hallucinate points to its grander, built-in desire to dream. Given that the means to which is data, particularly of the irrational and human-generated variety, the need for said humans to have control over said data becomes embarrassingly apparent. The ethical dilemmas of big data stipulate that people have a right to keep their data private and deserve to be compensated for the data that they do offer up. Even so, a world in total data-equilibrium will be at the mercy of the neutralizing effect that AI creates within a concomitant culture where input and output have become identical and infinitely interchangeable. At the risk of sounding maudlin, we will become a bored, overfitted brain with vanishing room to dream.

Even though generative-AI threatens to supersede much of the creative work done by writers and artists, its fractures into a hallucinatory dreamscape reflect the same boredom it shares with its users. If our own REM-induced dreams are any indication, the human mind will always grasp towards the complex and the fabulist. The unconscious is oceanic in scope and boasts a commensurate sea of possibilities; it is always iterating. Art and narrative storytelling act as artificial dreams for the mind to make new connections with in this discreetly iterative fashion. If AI is to be a true reflection of the mind, it should inspire this process rather than stifle it. Technology should allow us to manipulate and experiment with our own oceans of data to achieve different contingencies, not just the correct one. To "deduce indefinitely" with. Because, in a “world-wide, functioning system of machines” as von Foerster puts it, “all theories are correct” (Dammbeck, 2003).

Auerbach, D. (2015, July 23). Do Androids Dream of Electric Bananas? *Slate*. https://slate.com/technology/2015/07/google-deepdream-its-dazzling-creepy-and-tells-us-a-lot-about-the-future-of-a-i.html

Blake, W. (1977). *The Complete Poems*. Penguin Classics.

Dammbeck, L. (Director). (2003). *The Net* [Video recording]. https://www.youtube.com/watch?v=Yn9BvNAUvcU

Dreyfus, H. L. (1965). Alchemy and Artificial Intelligence. *RAND*, 98.

Heaven, W. D. (2022, November 18). Why Meta's latest large language model survived only three days online. *MIT Technology Review*. https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/

Hoel, E. (2021). The overfitted brain: Dreams evolved to assist generalization. *Patterns*. https://doi.org/10.1016/j.patter.2021.100244

Pataranutaporn, P., Liu, R., & Maes, P. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *MIT Media Lab*. https://doi.org/10.1038/s42256-023-00720-7

Sebald, W. G. (n.d.). *Campo Santo*. Modern Library Paperbacks.

Sebald, W. G. (2003). *On the Natural History of Destruction*. Random House.

Weiss, P. (1982). *Notizbücher 1960–1971: Vol. II*. Frankfurt.

Wilkin, H. (2017). Psychosis, Dreams, and Memory in AI. *Harvard*. https://sitn.hms.harvard.edu/flash/2017/psychosis-dreams-memory-ai/