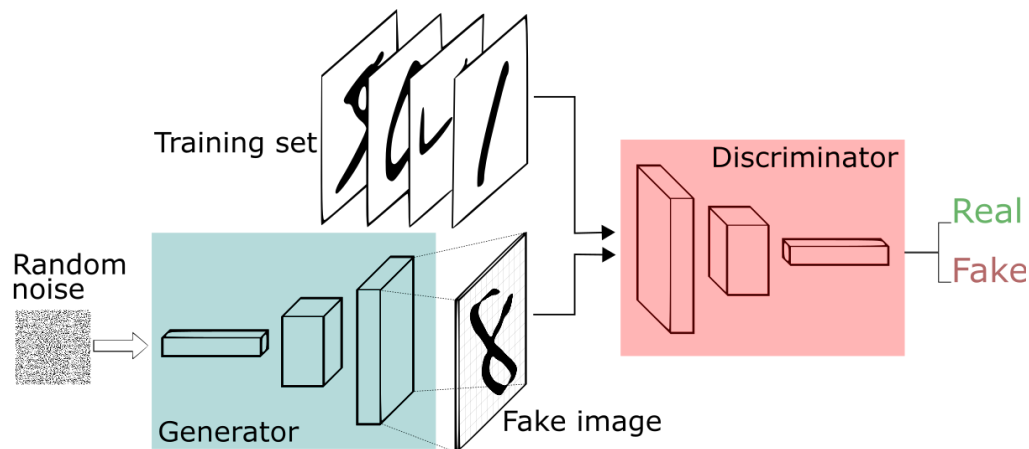


## Langland-Hassan summary

- Part of a larger book on creativity and imagination
- Generative adversarial networks



- Question: can a GAN be said to be creative?
  - They argue that human creative cognition may rely on similar process as how GANs are trained, and so we should not rule out the premise
- Human artists may receive “unsupervised training” by looking over many works, and learning regularities of the genre. Eventually, they become sufficiently adept as critics to improve their own work
  - How the artist comes up with an idea is a mystery to themselves, as they cannot provide clear reasons for why they did what they did
- However, they distinguish between ordinary creativity (combining concepts) and transformative creativity (new style, “revolutionary” works)
  - I don't think there's a clear delineation between these two things...
- Maybe an artist's creative process could be analogized as first generating/imagining some possible works in a genre, and then judging whether or not this work is considered to be good
- Parallels discussions around Midjourney, Stable Diffusion
  - E.g. is the model “copying” works in its training set, is it comparable to humans in ‘drawing influence from’ various other artists
- Not sure if the GAN analogy specifically is apt, but generally sympathetic to idea

Questions:

- What is bullshit?
- How is bullshit different from lying?
- Why is there so much bullshit in the world?

The term *bullshit* is used in a loose and vague manner, so it's hard to give a precise account of necessary and sufficient conditions for bullshit, but this essay is an attempt to provide some notion of an account.

### What is bullshit?

Max Black's definition of *humbug*: "deceptive misrepresentation, short of lying, especially by pretentious word or deed, of somebody's own thoughts, feelings, or attitudes"

- *Deceptive misrepresentation*: Does bullshit depend on state of mind (intent to deceive) or does it have some essential characteristic as an utterance?
- *Short of lying*: Is the difference between lying and bullshitting a matter of degree or kind?
- *Especially by pretentious word or deed*: Bullshit need not be motivated by pretense.
- *...somebody's own thoughts, feelings, or attitudes*: Black's account suggests that people want to deliberately misrepresent the state of their own mind to others. It is short of lying since the statement may not be false of the world, but false of the perpetrator's state of mind. Frankfurt thinks this falsity is not an essential character of bullshit. Also notes that bullshit can be in deed and not in word alone.

Fania Pascal's Wittgenstein anecdote: "I had my tonsils out and was in the Evelyn Nursing Home feeling sorry for myself. Wittgenstein called. I croaked: "I feel just like a dog that has been run over." He was disgusted: "You don't know what a dog that has been run over feels like.""

Assuming Wittgenstein wasn't playing along with the joke, what could his objection be?

- Pascal is not lying. At the very least, it is safe to presume a run over dog feels bad, as Pascal did.
- But Pascal has spoken of a specific feeling of which she can attest no knowledge. The statement is made without concern for what a run over dog actually feels like, and it's this mindlessness that might have upset Wittgenstein.
- When engaged in conversation, Wittgenstein works under the assumption that what Pascal says is carefully chosen to express her belief, but Pascal is carelessly violating this norm.

It is this carelessness – an *indifference to truth* – that Frankfurt identifies as the essence of bullshit. "Unlike plain lying, however, it is more especially a matter not of falsity but of fakery. This is what accounts for its nearness to bullshit. For the essence of bullshit is not that it is false but that it is phony. In order to appreciate this distinction, one must recognize that a fake or a phony need not be in any respect (apart from authenticity itself) inferior to the real thing. What is not genuine need not also be defective in some other way. It may be, after all, an exact copy. What is wrong with a counterfeit is not what it is like, but how it was made."

In some cases, this expectation is suspended to begin with, such as a *bull session*. This notion of bullshit extends to actions, where pointless activity is “bull” if it doesn’t serve the broader intent.

### **How is bullshit different from lying?**

Another point of difference between bullshit and lies is that someone who is lying is still concerned with the truth, and is contradicting it. Bullshitting involves no concern for the truth. “It is impossible for someone to lie unless he thinks he knows the truth. Producing bullshit requires no such conviction. A person who lies is thereby responding to the truth, and he is to that extent respectful of it. When an honest man speaks, he says only what he believes to be true; and for the liar, it is correspondingly indispensable that he considers his statements to be false. For the bullshitter, however, all these bets are off: he is neither on the side of the true nor on the side of the false.” This, Frankfurt argues, is a reason bullshit is a greater enemy of the truth than lies – lies acknowledge and reject the truth, while bullshit pays no attention to the truth at all.

### **Why is there so much bullshit in the world?**

Specifically, why so much bullshit now? Is it that the proportion of bullshit is the same, but as a result of how much communication there is in general, there is a greater absolute amount of bullshit?

In a democracy, a citizen is expected to have opinions about everything, whether or not they have knowledge of that thing. These opinions, given that they stem from ignorance of facts, are essentially bullshit.

Another reason for the proliferation of bullshit is the rejection of the existence of an objective reality. If the expectation is that what one says is not true of the world at large, but is a sincere representation of one’s beliefs, then what one says is disconnected from what is true, and hence essentially bullshit.

But Frankfurt argues that if we doubt the existence of a stable objective reality, then we must also doubt the existence of our own nature as something that we can be sincere about. So, attempts at such sincerity are themselves bullshit.

# ChatGPT Is a Blurry JPEG of the Web

TED CHIANG

*“So just how much use is a blurry JPEG, when you still have the original?”*

## The case of a Xerox photocopy

- Starts off with how lossy compression in general can make subtle errors, often at the cost of accuracy.
- Discusses how a photocopying machine printed the same area label for three rooms, when the labels were actually different for each room
- They use a lossy compression format, where the copier identifies similar-looking regions in an image, stores a single copy, and reuses that repeatedly during decompression
- Issue is not with the actual “blurriness”, but the fact that it is producing incorrect but highly plausible (here, readable) outputs
- *“File compression dances a two-step, gracefully encoding and decoding data into smaller forms. When files twirl back flawlessly, they've danced the lossless waltz, but if they return as approximations, they've stumbled into a lossy jig.” - summarized by ChatGPT*

## How does this analogy help understand the current state of LLMs?

- *“Hold onto your photocopiers, folks, because we're about to dive into a whirlwind of AI shenanigans that'll make your head spin faster than a buffering internet connection!”*
- Creates a hypothetical scenario where you are about to lose internet access forever
- You have 1% of the space needed to store data, but ~~you~~ (“OpenAI”) have unlimited compute and so design an algorithm that stores statistical regularities
- However, one can never get an exact match
- *“It's like looking at a Monet painting—up close, it's a mess, but step back and it looks pretty darn good. Sure, there's some loss of information, but like a master of disguise, ChatGPT dresses up the approximations in grammatical text, making them quite acceptable.”*
- Hallucinations are compression artifacts, but are plausible enough that identifying them requires comparing them against the originals (which means either the Web or our own knowledge of the world)
- Common method of compression is interpolation. ChatGPT does something similar by taking two points in lexical space and generating text between them (and is damn good at it)
- *“So buckle up and enjoy the ride, because with ChatGPT, you're in for one wild, pun-tastic adventure!”*

“Are large-language models like ChatGPT just lossy text-compression algorithms, or are they the ultimate smarty pants?”

- Given their current hype, it may sound dismissive to say so
- However, it's a useful corrective to the tendency to anthropomorphize LLMs
- *“Enter Marcus Hutter, AI researcher and quizmaster, offering cold hard cash for anyone who can smoosh Wikipedia into a tighter package than the last victor. He believes that, to reach peak compression, you've got to get chummy with the text—really understand it, like two peas in a pod.”*
- Imagine you have text files with just mathematical equations: best compression can probably be achieved by deriving the principles of arithmetic and writing code for a calculator.
- For text : If a compression algo knows that  $F=ma$ ; it can discard words from related physics documents and so on
- *“If a LLM has compiled a vast number of correlations between economic terms—so many that it can offer plausible responses to a wide variety of questions—should we say that it actually understands economic theory?”*
- Coming back to arithmetic : The model does extremely well when numbers have two digits, but accuracy drops to 10% for numbers with 5 digits. It doesn't simply memorize but has also been proven unable to derive the rules of arithmetic. **Then how does it get the right answers when it does? What is the role of chain-of-thought reasoning here?**

Given GPT-3's failure at a subject taught in elementary school, how can we explain the fact that it sometimes appears to perform well at writing college-level essays?

- Is it possible that, in areas outside addition and subtraction, statistical regularities in text actually do correspond to genuine knowledge of the real world?
- Proposes a human learning analogy to explain this: If ChatGPT were lossless, it would produce text verbatim, which isn't half as impressive as rephrasing words. Just like in humans, rote memorization is often not an indicator of genuine learning
- *“Welcome to the ChatGPT show, where we ponder the mysteries of AI that can write college-level essays but trip over arithmetic like a math-phobic toddler.”*

What (and what not) can these models be well suited for?

- Can LLMs take the place of traditional search engines?
  - Fabrication / hallucination is a huge problem here
  - By publishing LLM generated text on the web, it only becomes a blurrier version of itself.
  - When OpenAI trains the next GPT using data generated from a previous version, we can conclude that the analogy of lossy compression is not applicable
- Can large-language models help humans with the creation of original writing?
  - Can use them as an aid, but the author argues that using them to create original work is not good

- The effort spent in first creating and polishing un-original work is essential to eventually create original works of your own.
- It's only in the process of writing that you discover your original ideas.

*“So, do we really need a blurry jpeg when we've got the original? Unless we're about to lose the internet forever, maybe it's time to put down the AI photocopy machine and start writing our own masterpieces. The future of AI writing awaits, but for now, let's savor the joy of human creativity!”*

## Some fun with ChatGPT

**Prompt** : Could you please summarize this in a creative way that I can present in a philosophy class? Feel free to add subtle puns in between <insert 1st para>

*“Once upon a time, in the land of German construction, a curious case of the shapeshifting photocopier emerged. Three rooms, each distinct in size, awaited their duplication destiny. But alas, the enchanted Xerox machine had other plans.*

*With a mischievous twinkle, it transformed their identities, bestowing upon each a uniform area of 14.13 square meters. The perplexed workers, vexed by this puzzling metamorphosis, called upon the wise sorcerer of computer science, David Kriesel, to unravel the conundrum.*

*Kriesel, a sage in digital alchemy, cast his gaze upon the contraption, revealing its arcane secrets. The ancient xerographic process, once the bedrock of duplication, had given way to a digital scanning ritual and the printing of its image offspring. The plot thickened as Kriesel unveiled the insidious role of compression, a space-saving enchantment that inadvertently conjured the illusion of uniformity. And so, the tale of the shapeshifting photocopier was laid bare—a testament to the unforeseen consequences of digital wizardry and a reminder that, in the realm of image sorcery, not all rectangles are created equal.”*

**Prompt** : Could you not get into your storytelling mode and simply summarize the following? You can add subtle puns if you'd like <insert 2nd para>

*“File compression dances a two-step, gracefully encoding and decoding data into smaller forms. When files twirl back flawlessly, they've danced the lossless waltz, but if they return as approximations, they've stumbled into a lossy jig.*

*Text files and computer programs prefer to tango with lossless compression, as a single misstep can spell disaster. Meanwhile, photos, audio, and video are content to sashay with lossy compression, where absolute accuracy isn't the belle of the ball.*

*Most of the time, we don't bat an eye at minor imperfections in our pictures, tunes, or films. However, when files are squeezed into their dancing shoes too tightly, we can't help but notice the awkward compression artifacts—jpeg's fuzzy pixels or the tinny sound of low-bit-rate MP3s—stumbling across the digital dance floor.”*