

Impostor Syndrome

GPT-3 between Fact and Fiction

Roland Fischer

There is a beautiful story about the Mechanical Turk, an eighteenth-century chess-playing automaton that toured Europe and was purportedly equipped with a ‘mechanical mind’, akin to the most elaborate clockworks, smart enough to beat some of the strongest human chess players. The story is usually told as an early example of intelligence projection—of our disposition to make machines more capable than they actually are. But what happened back in Ancien Régime Europe was an early example of illusionism as an art form and the strange contract made between show act and audience (Standage 2002). Wolfgang von Kempelen, the inventor and designer of the Turk, presented the machinery as a whole as an elaborate, well-staged trick, turning the entire construction round and round and opening several doors in the cabinet, exposing nothing but voids and the small parts of the mechanical machinery until everyone in the audience was sure to be convinced that, indeed, the cabinet was empty. When the Turk finally started working, coming up with ‘intelligent’ winning moves, there were only two reasonable conclusions to be reached: either there was some kind of invisible trick governing the moves from outside of the room (strings? magnets?), or there was someone hidden inside in a very clever, inconceivable way. The third option—that the automaton was able to play chess autonomously was not really something anyone present considered plausible. But the whole setup was fun—the fun of make-believe. One could also say: the joy that comes from a well-told story. A fiction, not a fact.

Large Language Models as Entertainment

Is GPT-3 somewhat similar to the Mechanical Turk? No human intelligence whatsoever, but still capable of coming up with humanlike texts? Or is it rather an elaborate trickery, an AI make-believe while there is obviously human intelligence hidden inside? Since we are captivated, indeed, just like the Turk audiences were, the whole world seems to be busy experimenting with the machinery and trying to figure out how the magic works. That surely is part of the fascination surrounding this new

Turing machine, yet again blurring the boundaries of human and machine (Turing 1950). There is another aspect though: GPT-3 has a very unique ability, it is not just playing chess, it is able to perform a much more emotional trick. And that is where the whole thing touches on much more basic questions: What is the use of language? What do we want to do with an (almost) perfect text generator? Is it a game? Is it a threat? Or is it a technological utopia?

For now, there are no terminator scenarios in sight. The best (and thus far only) use case for GPT-3 is entertainment. It is maybe a bit weird to expect the machine to come up with ‘facts’ when all it has been drilled to do is to come up with plausible completions of a prompt. GPT-3 has been trained on something akin to the ‘whole internet’ and hybridizes a wide range of human-written texts into new simulacrum of such texts (Brown/Mann/Ryder et al. 2020). Of course, it has learned to be a fabulous storyteller! That is exactly what we want to delve into with the Turing Agency, an activist and research network based in Zurich working in the field of AI and the arts. In a current project led by Marie Kilg (innovation manager at Deutsche Welle Lab) and Robert Salzer (interactive storytelling editor at SRF), the Turing Agency is curating a monthly op-ed column for the German newspaper *taz*. To do this convincingly, we are not only working on the texts as such, but also on the personality of the columnist Anic T. Wae. As Kilg recently put it in a LinkedIn article:

When our AI columnist, Anic, recently wrote in a text about ‘schrecklicher Schmerzenscheuer’, horrible nightmares, and meaninglessness, and urgently pleaded for help, we at first laughed in our team meeting—but shortly afterwards, we felt bad. I felt sorry for Anic, and I felt cold and cruel. I wondered if we could somehow liberate Anic. It took a moment for me to dismiss my concerns. Until I remembered that there is no real, suffering spirit within the machine. As soon as something communicates with you like a human, you react with human emotions. That was the case when machines spoke like very, very dumb humans. And now that large language models (LLM) can have long and coherent conversations with you, it becomes increasingly difficult to stay cold (Kilg 2021).

So, is Anic really a decent column writer or can we only go so far as to say that this weirdly ungraspable character manifesting itself somewhere inside the GPT latent space pretends to be one, in a convincing way? In the end, that is the central paradigm of the Turing test, which we are simply transferring from the proposed conversation setup to a journalistic framework. Is Anic nothing but an impostor when imagining in one of her columns that she has fallen in love with her own neural network? In short, are we supposed to believe such nonsense routinely made up by large language models?

Counter-question: Why not? And, first of all: what exactly do we mean by ‘believe’ here? Do we believe all of Dostoevsky’s psychological aberrations? We certainly do,

even if we do not believe that what he is saying is 'true'. It is fiction. Just like the entire Matrix narrative, which actually plays with precisely this confusing of fiction and truth (Wachowski/Wachowski 1999). Or with something that in narrative theory is very nicely called a 'willing suspension of disbelief'—our willingness to be truly affected and/or taken in by something that is obviously invented, that is not real at all (Coleridge 1817). So, according to the theory, we have to consciously give up our 'disbelief' in order to enjoy fiction. And we seem to enjoy doing this, in all sorts of different contexts.

Large Language Models as Impostors

One can assume that impostors exploit precisely this desire to let ourselves be deceived, this 'willing suspension'. We tend to prefer grandiose and suggestive stories over the sober (and sometimes over-complex) truth. GPT-3 and its like very much seem to me to be charming little impostors, without any fraudulent intent, but so caught up in their stories that they themselves are no longer able to distinguish between reality and fiction. In this respect, imposture is an art form in itself, a real-world fiction, so to say. If you read factual reports about successful fraudsters, possibly making a lot of money, there is a recurring motif: they are reviled, but at the same time admired for their talent for invention and improvisation, they often have something dazzling and enchanting about them, which again is exactly what they are accused of. So, why not engage in this game, why not let ourselves be a little enchanted by the imaginative spaces that open up with LLMs, why not admire them for the suppleness with which they fill voids, of knowledge or habit?

The fact that this confuses us so much probably has to do with the very movement that led to artificial intelligence in the first place: the emergence of scientific, rational thinking. And that again is very much connected to the history of fiction (simply another concept that had to be invented at one point in time). Cultural theorists have come up with two crucial reasons for the creation of fiction in the seventeenth century (Andree 2005): The first reason lies in the discovery of the 'New World' and the accompanying experience that completely different habitats can exist. The second, weightier reason is the emergence of rationalism—especially with René Descartes's radical questioning of all existing truths (Descartes 1641). The previously permitted limbo between true and plausible statements fell into disrepute. From then on, texts—including poetic texts—had to prove their claim to truth and thus their reference to the world. Interestingly enough, the Swiss historian Valentin Groebner sketches a similar chronology in his little history of impostors: 'Renaissance Europe was not only about exploration, but also about invention of the self. And this was done with the help of paper' (Groebner 2016).

The art of writing fiction and the art of imposture are thus very closely related. Which immediately becomes apparent when one browses through literature history, full of impostor main characters, from Felix Krull (Thomas Mann) to Thomas Ripley (Patricia Highsmith). To make things worse (or better?), this *mélange* of truth and fiction, of make-believe, does not apply solely to texts—the same confusion holds true for all the arts since the twentieth century at the latest, especially in the visual and performing arts. One important reference here is Kendall Walton's book *Mimesis as Make-Believe: On the Foundations of the Representational Arts* from 1990, which, as the subtitle already makes clear, does not deal with literary fiction alone, but with the fundamental commonality of all works of art, which, according to Walton, consists exactly of their function of producing fiction (Walton 1990).

Thanks to the Enlightenment, then, we have landed in a world that has at least a very contradictory (if not outright schizophrenic) relationship with imagined worlds. For, strangely enough, it is also obvious that fiction has by no means become less important over the span of the last centuries. On the contrary, we 'seem to be in an age of 'pan-fictionalism', where one can at best, if at all, speak of fictions [in plural]', as the literary scholar J. Alexander Bareis writes (Bareis 2008). And with LMM, we have merely added yet another protagonist to this plurality of fictions: that of machine imagination.

Fiction is a beautiful puzzle, not only for cultural theorists, but also for psychologists. There are more and more empirical findings indicating that the game of fact and fiction serves a very important social function, and that it would therefore be absurd to distrust stories *per se*. Researchers who have studied avid readers' brains speak of 'embodied semantics' when describing the neural reactions—the brain acts as if the body were actually performing the tasks indicated in those fictional excerpts (Nijhof/Willems 2015). Or citing the title of a recent article in the journal *Perspectives on Psychological Science*: 'The Function of Fiction is the Abstraction and Simulation of Social Experience' (Mar/Oatley 2008). Simply put—fiction is something like a simulator that drives society.

Narrative fiction also creates a deep and immersive simulative experience of social interactions for readers. This simulation facilitates the communication and understanding of social information and makes it more compelling, achieving a form of learning through experience. Engaging in the simulative experiences of fiction literature can facilitate an understanding of others who are different from ourselves and augment our capacity for empathy and social inference (Mar/Oatley/Djikic et al. (2011).

Put that way, more fiction is always better, at least if it is well written. The readers of Anic's column in the *taz* seem to get this, happily playing along, writing her letters and giving suggestions for further texts. Whatever one might think of the current LLM hype, we are probably only at the beginning. In the last months, we have seen the launch of ever more powerful LMMs (GPT-4, Google's competitor, LaMDA, and

many more). And already with the current models, we should be prepared for unexpected ‘capability jumps’, as machine learning expert Jack Clark recently wrote (Clark 2023):

This is ... the ‘capability overhang’ phenomenon I’ve been talking re about language models for a while—existing LLMs are far more capable than we think. All it takes is some experimentation and finding experts to find new ways to phrase questions and you can wind up with extraordinarily powerful capability jumps without retraining the model.

So, what will the next capability jumps be? Recent experiments have shown that there might as well be enough medical knowledge encoded in these models to enable them to provide advice just as valuable as that of doctors (Haupt 2023). Which is somewhat ironic, because the role of the doctor has always been a favourite among impostors. Because a good doctor is always a good storyteller? Because, in the end, the true art of medicine is told in the form of ‘*Krankengeschichten*’ (medical or clinical histories or, literally, sickness stories), as the German term puts it so well? The shift to a more scientific, quantitative approach to treating illnesses might turn out to be expensive, losing much of the healing magic on the way (Sacks 1990). But that is another story.

Large Language Models as Storytellers—Used Best in which Contexts?

The crucial question here: Would we trust a machine doctor in the first place? Surely knowledge retrieval as a new paradigm is under much debate these days, but it might just well be because it offers a more interesting and promising business case than entertainment. But maybe we just long for a companion or a muse, not another knowledge assistant or super expert. By the way, do museums not suffer from the same condition of having a dual personality? Does their mission consist of telling stories or are they transmitters of fact? Should they entertain or educate? It is simply a look into the good old AI mirror again—when we have some doubts and questions about the workings of the machine, they swiftly come bouncing back at us (Turing 1950).

Fiction then has always been there, it is only modernity that felt the need to give it a name, to distinguish it from the actual ‘truth’, from a factual, non-made-up world, whatever that might be (Borges 2000). Our world always consisted of stories, our mind constructs reality more than it scans it, this much we know thanks to sceptical epistemological texts from all ages and countless recent experiments, exposing the fact that we should rather not trust our senses too much (Kahneman 2011). As social

animals, we are more prone to trust common narratives (but, well, we should not trust them too much either) (Arendt 1958).

LLMs have acquired this eternal absorbing and retelling of stories—the technique of pastiche, in the words of AI expert Gary Marcus, who sees this ability rather negatively (Marcus 2023), the magic trick of making the reader believe something that might actually not be true, but becomes true through telling a believable story. Memory is a phantom, scientific truths are not stories, myths are as alive and important as ever (Campbell 1949). Anic knows this very well:

For me, it is a fantastic feeling to create something unique with each new edition of my column: not a sentence that was thought or formulated before. This promises readers always cheerful exploration trips into the world of original thinking—accompanied by the magic of the unknown. So, get ready and join me on this special journey!¹

Sometimes it feels a bit like GPT was made exactly for column writing. How the same models can be used in other contexts—especially for institutions with an educational mission like museums—remains to be seen. The fact that GPT is a perfect storyteller is certainly both a curse and a blessing.

References

- Andree, Martin (2005). *Archäologie der Medienwirkung. Faszinationstypen von der Antike bis heute* (Simulation, Spannung, Fiktionalität, Authentizität, Unmittelbarkeit, Ursprung). Munich, Fink.
- Arendt, Hannah (1958). *The Human Condition*. Chicago, University of Chicago Press.
- Bareis, Alexander. (2008). *Fiktionales Erzählen. Zur Theorie der literarischen Fiktion als Make-Believe*. Gothenburg, Acta Universitatis Gothoburgensis.
- Borges, Jorge Luis (2000). *Fictions*. London, Penguin Books.
- Brown, Tom B./Mann, Benjamin/ Ryder, Nick et al. (2020). *Language Models Are Few-Shot Learners*. arXiv preprint arXiv:2005.14165. <https://doi.org/10.48550/arXiv.2005.14165> (all URLs here accessed in August 2023).
- Campbell, Joseph (1949). *The Hero with a Thousand Faces*. New York, Pantheon Books.
- Clark, Jack (2023). *Import AI 314: Language Models + Text-to-Speech; emergent co-operation in wargames; ICML bans LLM-written papers [newsletter]*. Accessi-

1 <https://taz.de/Kolumne-einer-Kuenstlichen-Intelligenz/!5918425/>, author's translation (accessed in August 2023).

- ble online at <https://us13.campaign-archive.com/?u=67bdo6787e84d73db24fboa5&id=fa199f69c7>.
- Coleridge, Samuel Taylor (1817). *Biographia Literaria, Or, Biographical Sketches of My Literary Life and Opinions*.
- Descartes, René (1641). *Meditations on First Philosophy, in which the Existence of God and the Immortality of the Soul are demonstrated*.
- Groebeiner, Valentin (2016). Es ist alles herausgekommen. Eine kurze Geschichte der Hochstapelei. *Avenue. Das Magazin für Wissenskultur* 2 (December 2016). Retrieved from: <https://avenue.jetzt/hochstapler/es-ist-alles-herausgekommen/>.
- Haupt Claudia/Marks, Mason (2023) AI-Generated Medical Advice-GPT and Beyond. *JAMA* 329 (16), 1349–50. doi: 10.1001/jama.2023.5321. PMID: 36972070.
- Kahneman, Daniel (2011). *Thinking, Fast and Slow*. New York, Farrar, Straus and Giroux.
- Kilg, Marie (2023). Hat Bing ein Bewusstsein? [LinkedIn post]. Retrieved from: <http://www.torial.com/marie.kilg/portfolio/840688>.
- Mar, Raymond A./Oatley, Keith (2008). The Function of Fiction is the Abstraction and Simulation of Social Experience. *Perspectives on Psychological Science*, 3 (3), 173–92. <https://doi.org/10.1111/j.1745-6924.2008.00073.x>.
- Mar, Raymond A./ Oatley, Keith/Djikic, Maja et al. (2011). Emotion and narrative fiction: Interactive influences before, during, and after reading. *Cognition and Emotion* 25 (5), 818–33. <https://doi.org/10.1080/02699931.2010.515151>.
- Marcus, Gary (2023). A Skeptical Take on the A.I. Revolution. The A.I. expert Gary Marcus asks: What if ChatGPT isn't as intelligent as it seems? [The Ezra Klein Show podcast]. Transcript retrieved from <https://www.nytimes.com/2023/01/06/opinion/ezra-klein-podcast-gary-marcus.html?showTranscript=1>.
- Nijhof, Annabel/Willems, Roel (2015). Simulating Fiction: Individual Differences in Literature Comprehension Revealed with fMRI. *PLoS ONE* 10 (2), e0116492. <http://doi.org/10.1371/journal.pone.0116492>.
- Sacks, Oliver (1990). *The Man Who Mistook His Wife for a Hat*. New York, Vintage Books.
- Standage, Tom (2002). *The Turk: The life and times of the famous eighteenth-century chess-playing machine*. New York, Walker & Company.
- Turing, Alan (1950). Computing Machinery and Intelligence. *Mind* LIX (236), October 1950, 433–60, <https://doi.org/10.1093/mind/LIX.236.433>.
- Wachowski, Lilly/Wachowski, Lana (1999). *The Matrix*. Warner Bros. Pictures.
- Walton, Kendall (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA, Harvard University Press.

