ChatGPT and Generative AI, between Farce and Tragedy

SIMONE VANNUCCINI - GREDEG – UNIVERSITÉ CÔTE D'AZUR - CNRS

I do not know whether Karl Marx, when he wrote that history repeats "the first time as tragedy, the second time as farce" in "The Eighteenth Brumaire of Louis Bonaparte", also considered the possibility of the dynamics moving in the opposite direction, from farce to tragedy. However, the current trends revolving around the latest artificial intelligence (AI) models might be a good case to test the hypothesis. Since a few months, the world has been taken by storm by the latest iteration of what has been labelled "generative" AI, OpenAI's chatbot ChatGPT. ChatGPT is not the only generative AI in town, but it is the one that achieved record widespread diffusion in a very short time. Rivers of real and virtual ink have already been spilled to dissect its nature, use cases, and impact on industry organisation and society at large – and, of course, also to feed speculation, hype, and prophecies on the always imminent next revolution to come. At the same time, the success of the technology contributed to kick-start a new wave of interest in AI, snoozing to the future the grounded fears that a new AI "Winter" – a phase of shift of attention and funding away from AI – was looming behind the corner.
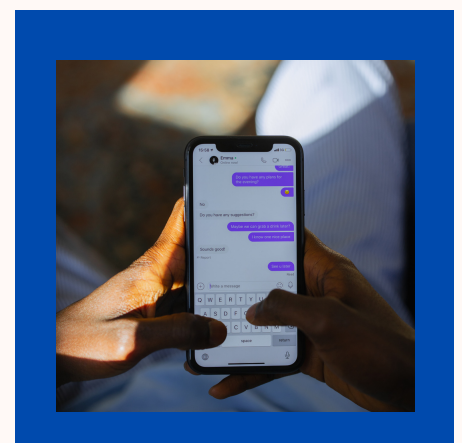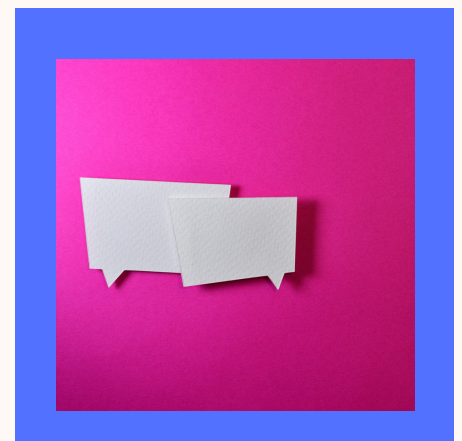
How should we think about technologies of the likes of ChatGPT? Let me summarise my take on the matter before proceeding: generative AI is a powerful new tool, and yet it is just a tool. As it happens with every tool, from forks and knives (double-stranded to the emergence of "good manners" in the Middle-Ages) to the ball-point pen, from spreadsheet software to tunnelling microscopes, we use and shape it, but it also shapes and "uses" us. This specific tool will not reshape us and our societal interactions in the same radical manner as electricity or computing did. Rather, it will find its well-defined place amongst human activities as soon as the hype-tide calms down and we learn to use AI powered agents on a day-to-day basis, for example by adapting the way we assess students in higher education.

In a nutshell, the very nature of ChatGPT is not as conducive of profound, pervasive transformations as it seems: rather, the technology is a particular piece of equipment – in this case digital equipment. This means that as you do not cite your laptop as a co-author of your essays, you should not cite ChatGPT either.

Instead, what is problematic is the risk of attaching to this piece of digital equipment capabilities that are different from those it really has, and to have economic actors building business models to start extracting value from it without precaution.

Getting back to Marx's quote, it is not the first time that we interact with "intelligent" chatbots: in 2016, Microsoft's Tay became the posterchild of history as farce – and as so it had to be retired in a rush after a short while, having turned into a bleak mapping of the racist and misogynist Twitter. ChatGPT, a chatbot based on different working principles and displaying more general capabilities, might become the epitome of history as tragedy, given that the stakes are much higher now, and bigger are the possible damages the technology can produce. Consider the issue in terms of opportunity costs: while it was easy for Microsoft to shut down Tay, the decision to take ChatGPT back from the wild in which OpenAI set it free entails very large costs, given that ChatGPT has rapidly become the cornerstone of the most recent Microsoft attempt to fight back in the search engine war (read: the fight for market shares in the digital advertising market) against the dominant Google. Embedding ChatGPT-style modules in existing online services is a much more far-reaching strategy. Hence, if this technology is so key to a digital giant business model, then its drawbacks are not anymore embarrassing issues to forget quickly, but collateral damage. And the way to deal with collateral damage is to do damage control and patchwork fixes, rather than to retreat the product from the market, regardless of the magnitude of the negative impacts it produces.

To understand the tragedy potential of (as well as the opportunities linked to) ChatGPT, a little explanation is in order. ChatGPT is a specific type of AI – or, better, of AI "solution" – which, in turn, builds on a specific class of AI models, those algorithms that process text/language data and that are called foundation or large language models (LLMs). In particular, the -GPT part of ChatGPT stands for Generative Pre-trained Transformer, meaning that the "engine" that processes the data and that is powering the AI solution is a so-called Transformer, a particular architecture of the algorithm (funnily, first introduced by Google Brain researchers). There are other architectures available in the AI world (e.g., the so-called "diffusion" models used by some of the image generation AI solutions), but the Transformer has quickly risen to the role of quasi-dominant design in the field. This algorithm is pre-trained on large corpuses of textual data, so that it acquires the probability distribution of co-occurrence of words in human language and, as a consequence, can replicate or imitate it. With such distribution at hand, the system can be fine-tuned on domain specific data, allowing to create product variants (e.g., ChatGPT-like products that answer biomedical or chemistry questions already exist). The Chat- part of ChatGPT relates to the fact that the AI solution has a user-friendly interface that can be queried from the web with all sorts of text-based prompts. The interface allows for easy interaction with the AI and can be potentially embedded in an app or integrated within different websites, creating room to offer access in the form of "chatbot-as-as-service".

The success of ChatGPT stands precisely in the combination of the two components making up its name: a powerful predictive model of language, and an easy-to-use interface – so practical that a very novel field of "prompt engineering" has born to explore and trial the best ways to dialogue with the chatbot and to, well, engineer the direction of its answers. In sum, ChatGPT is an identifiable, well-bounded product (service), a recognisable tool that can be packaged, wrapped, embedded in other products and services, and sold (i.e., per access). From an economic and strategy viewpoint, this is the key feature of interest we should focus on, because it tells us how the ChatGPT and similar solutions can generate and capture value. It explains why Google – despite having introduced the computation engine behind all recent language models – still lags (and recently failed the launch of) final-users-oriented product commercialisation which, in medium term, translates into monetisation. As Azeem Azhar **put it**: *"interacting with a sophisticated chatbot is about an extended discovery process where you, to some extent, rely on the latent connections that systems make for you ... It's a new type of use, a new paradigm for interfaces."*

In sum, the technology is (relatively) replicable, while the interface is proprietary, as it is the smooth, seamless experience of interacting with the product. Value is created at the intersection of this new paradigm of interfaces with a powerful supporting AI language engine.

A powerful predictive model of language

All this, by itself, does not make the case for a possible tragedy – AI+interface just equals a new tool for humans to play with. A radical tool, surely, as it is progressively injected in many activities that, in the world pre-Chat-GPT, we conducted in rather different ways: search, summary, impromptu writing; however, still a tool. The problems emerge when the product is elected acritically to the role of hen with golden eggs because of its expected economic returns. AI systems are probabilistic systems, so they cannot expected to be error free – they can be likely correct, but not certainly correct. Let's focus on the case of search engines:

As Gary Marcus **makes clear**, *"(Traditional) search engines are databases, organized collections of data that can be stored, updated, and retrieved at will. (Traditional) search engines are indexes [...] Large language models do something very different: they are not databases; they are text predictors, turbocharged versions of autocomplete. Fundamentally, what they learn are relationships between bits of text, like words, phrases, even whole sentences. And they use those relationships to predict other bits of text. And then they do something almost magical: they paraphrase those bits of texts, almost like a thesaurus but much better. they are not databases; they are text predictors, turbocharged versions of autocomplete. Fundamentally, what they learn are relationships between bits of text, like words, phrases, even whole sentences. And they use those relationships to predict other bits of text. And then they do something almost magical: they paraphrase those bits of texts, almost like a thesaurus but much better. But as they do so, as they glom stuff together, something often gets lost in translation: which bits of text do and do not truly belong together."*

The probabilistic nature of AI language models is not an issue per se – rather, it becomes an issue when these solutions are used to search for complex information and facts that need to be correctly identified; in other words, in activities with high-stake loss functions, where we can't tolerate statistical error.

The ones outlined above are supply-side issues of the technology. Whether a tragedy unfolds depends on the demand side of the story: do users prefer an ok-ish search experience (i.e. one that might require additional, follow-on searches, back and forth to the search bar, link clicking, opening new tabs, etc.) with low risk of mistakes, or a seamless experience with high(er) risk of mistakes? I dare make a prediction: as humans are very often attracted to the most natural, economising/efficient, and integrated forms of communication (see the success of Emojis and memes as condensed, informationally-bounded, natural-feeling conveyors of meaning), AI interfaces have a pretty good chance to thrive – again in limited, though important because identity-defining – domains. ChatGPT is a product built on exceptional technological advances in AI techniques and models, and it offers a really novel way for users to play around with contents in the Web. It is an enabling tool: as it lowers the cost of engaging in interactions with information, we will see (and already see) a blooming of uses, including some very creative and some very malicious ones. But in terms of quality of its outcomes, due to its probabilistic nature prone to mistakes, it is yet a tool unfit to replace completely more structured ways to search for information. This does not mean it won't.

We know very well that technical superiority often does not stop inferior alternatives to thrive. If ChatGPT and similar solutions do succeed to become a new de-facto standard for interacting with the information space we call the Internet, it means that usability and good interfaces have lured us into a lower-level equilibrium. As it is having been very aptly described online, what ChatGPT outputs is essentially automated mansplaining, or "mansplaining-as-a-service": generic and lengthy "lectures" on any topic, produced with utmost confidence and ignoring the level of expertise of the prompter. This does not seem to be a problem – rather an issue of damage control – for the companies that started to integrate the solution in their existing services (read: Bing).

And the weight of AI chatbots limitations will be played down and swept under the carpet – in absence of a chatbot-backlash on the user side – the more the competition between different AI solutions offered by different providers will begin to bite. However, a damage that is only collateral for AI companies can be severe for the society. Just think of the exploitation of precarious digital labour contracted to label and moderate the text corpuses necessary to keep the AI models running, or to the spread of fake news, now with a more subtle a tricky twist made of linear combinations of correct and incorrect pieces of information offered by AI chatbots in naturally feeling, interaction-inducing, human-like dialogues.
A possible way to address the looming tragedy is to bring back the product "in vitro" – namely retarding distribution

and commercialisation while research and scientific advances proceed. However, this won't change the probabilistic nature of AI language models. An alternative is to put emphasis on LLMs that are open source and produced bottom-up by diverse communities. For these type of alternative "business" models of AI production to succeed, the magical mist around AI solutions needs to rarefy quickly, in order to give all of us a clear view on the incentives and constraints as well as market and non-market forces at work. If there is a domain where we should welcome probabilistic outcomes it is the unfolding of the future: the tragedy of AI systems following the farce might certainly realise, but it does not have to – it is completely up to us.