EIP

The Deep Non-Thinker

Lock up your authors! ChatGPT is here! And it's coming for your IP!

The internet is awash with content written about, or written by, the new chat-bot in town, ChatGPT. But what is ChatGPT? And what does it mean for the IP world?

ChatGPT, along with its geeky sibling Codex, artistic cousins Midjourney and Stable Diffusion, and a host of others already or soon to be released, are examples of "generative AI", whose emergence has the potential to fundamentally change the way we interact with technology. Software companies, desperate to cash in on the promise of increased productivity and newly terrified of becoming obsolete, are racing to integrate these new tools into their products. Others warn that the sudden and widespread deployment of generative AI could lead to an onslaught of dangerous consequences, including propagation of misinformation and fraud. Many are worried about the potential of generative AI to dilute and devalue human work, and to replace human jobs completely.

In this first of a series of three articles, I will explain what generative AI is and set out the strengths and limitations of the technology. In subsequent articles, I will discuss how generative AI may impact those creating and protecting IP.

What is generative AI?

Artificial intelligence (AI) refers to computer software that learns how to perform tasks by finding patterns in data, rather than being programmed by humans. Generative AI is a particular flavour of AI in which the computer generates new data that resembles data from which it has learned, such as text, audio, video or images.

In recent years, a few landmark innovations, along with increasing availability of data and sufficiently powerful computers, have provided the necessary conditions for generative AI to flourish into the headline-grabbing behemoth it is today, with ChatGPT placing the technology firmly in the public eye. By now, most of us have had a chance to marvel at the

amazing feats of generative AI, whether it is conjuring up a life-like image of an astronaut riding a horse, or explaining a mathematical algorithm in the style of a New York mobster.

How does it do that?

AI uses mathematical models to represent patterns in data. The model of choice for most AI applications is the (artificial) neural network, which specifies a sequence of mathematical operations that can be visualised as layers of artificial neurons and connections, loosely analogous to the physical components of a real brain.

As with a real brain, a neural network's ability to perform a task depends on the configuration of neurons and connections in the network, the quality and volume of data to which it has been exposed, and the way in which it has been taught. Generally speaking, the bigger (more neurons) and deeper (more layers) the network, the greater its capacity to learn – in theory, a big enough neural network can learn to represent any relationship that may be present in a set of data. But size comes at a cost: the bigger the network, the greater its appetite for data. And neural networks for generative AI can be ravenous indeed. The GPT-4 network behind the latest version of ChatGPT, for example, has of the order of a trillion neurons, and the cost of resources needed for its training likely ran into tens of millions of dollars as it was fed terabytes of text data scraped from the internet.

Size is not everything. The skill of an AI researcher lies in designing the configuration (architecture) of a network's neurons and connections to encourage it to find patterns in data, whilst keeping its hunger levels in check. Architectures are incrementally tweaked and improved all the time, but every so often there is a landmark innovation which results in a technological step change. For language-generating models (aka large language models, LLMs), this step change came from the invention of the "transformer" (the "T" in GPT), which is a neural network that uses "attention layers" to identify context-specific relationships between words (tokens) and their positions in passages of text (sequences). Things like New York's organised criminals frequently ending statements with a question, see? Attention layers provide the model with awareness of grammatical and semantic structures, and this is part of the magic that enables ChatGPT's musings to be consistently coherent and plausible. A much older, but analogous, invention is the convolutional neural network, which uses convolutional layers to identify spatial relationships between features in images, such as noticing that a person riding a horse always sits on the horse's back, rather than hovering above its head.

But these innovations alone don't explain the uncanny and sometimes eerily human-like outputs of ChatGPT. For that, an additional helping of secret sauce is needed, and that is the injection of randomness. Most generative AI models are not deterministic, and their

outputs vary randomly for a given input. By adding randomness during training, the model learns to handle unexpected variations, and accordingly to generalise more

effectively to situations it has not seen before. As a concrete example, ChatGPT predicts a most likely next word in dependence on its prompt and all of the words it has already typed, but with a small built-in probability of predicting a less likely word. This unpredictability is key to the model's training, with the bonus effect of making ChatGPT behave more like a human and less like a robot.

What's the catch?

A plausible output is not necessarily a good output. Generative AI sees only in patterns, and lacks any deeper understanding of its inputs or outputs. Although the reasoning ability of generative AI language models is continually improving, for now they still fail at relatively simple logic puzzles. An illuminating example is ChatGPT insisting that two kilograms of bricks weighs the same as one kilogram of feathers: the model has recognised the set-up of a common trick question, but erroneously generalised its clever-clogs answer to a different set of facts. AI practitioners would classify such behaviour as over-fitting. For similar reasons, a generative AI model for writing computer code is not guaranteed to produce the most logical coding solutions, only those similar to what it has seen in its training data. Generative AI is ultimately a parrot, not a deep thinker.

Another problem is that a generative AI model's memory of its training data is stored implicitly via the connections between its neurons, and as any forgetful human can tell you, this is not a reliable way of accurately storing information. Accordingly, the AI's internal representation of the training data is somewhat fuzzy – a blurry snapshot of the internet from when the model was trained. Any content created by such a model is dependent on this internal representation, as well as any data fed into it at execution time (e.g. a prompt/document to be summarised/image to be modified/random noise). The extent to which the model has inserted its own knowledge is not always evident, and this imports an inherent lack of reliability into any content it creates. In view of these consideration, perhaps the greatest peril of the current crop of generative AI models is that despite their shortcomings, they often present their answers with unflappable confidence.

The final consideration is the way in which generative AI models are trained, and who is the teacher. AI models are trained to maximise some sort of score or objective. This may be uncontroversial if the objective simply measures how faithfully the model reproduces patterns in data. However, to get ChatGPT to answer questions in a manner that is appealing to humans, a more complex objective, relying on human feedback, is needed. The humans designing the objective and providing the feedback wield great power over

the eventual outputs of the model. And these humans ultimately serve the interests of the commercial entity funding the process. ChatGPT's creator, OpenAI, is the first to admit this, but the potential centralisation of power to Big Tech's biggest guns has alarm bells ringing. On the other hand, a flurry of research activity in the open source software community has already succeeded in replicating and scaling down much of the functionality of generative AI models to the extent that they can be trained and run on a laptop, which may go some way to democratising access to the technology. Irrespective of who wins the race for generative AI supremacy, regulation will play an important role in keeping the balance of power and responsibility in check.

Conclusion

Generative AI is a powerful technology that is already transforming the way humans work and live. It is an illogical but persuasive beast, and its safe use demands a great deal of critical thinking, coupled with an awareness of its limitations and biases. The current generation of generative AI tools have a dangerous tendency to confidently deliver fabrications as truth – surely their most human-like quality of all.

In the next article in this series, I will look at the first touchpoint between generative AI and IP, which is: do generative AI models copy the content on which they are trained? Are the content authors' IP rights infringed in this process, and how should the law deal with this?