Environment and Planning B: Editorial: 50, Issue 3, 2023

A New Kind of Search

ChatGPT (2022) was first launched on 22 November 2022, but in only 3 months, it has attracted enormous interest. Reportedly, there are now over 100 millions users of the latest system (https://increditools.com/chatgpt-statistics/) that comes from the company *OpenAI*. It is hard to know if any software has had such an impact over so short a period of time. *ChatGPT* means **Chat G**enerative **P**re-trained **T**ransformer. It is software in which any user can 'chat' or communicate using digital media, usually on the Internet, primarily to engage in search which can be quite a creative process. These systems access a much wider range of data than the traditional search engines such as Google which mainly use Page Rank algorithms. Various kinds of machine learning are employed in these Chatbots, and to an extent, search is not their only function.

So what is all this hype about that is forcing us to figure out if the software is of any use to research and other activities in our own field (or anybody else's for that matter). Search, as we have argued in these editorials almost 20 years ago (Batty, 2005, 2006), is one of the key elements of the digital revolution. Essentially it now appears that *ChatGPT* is perhaps the most sophisticated search engine available and it enables us to use new forms of what are called 'large language models' to search much more effectively than we have ever been able to do in the past, so its originators argue. Large language models are autoregressive structures involving probability distributions that define sequences of words which are trained on a vast corpus of data upon which any search request can be based. Such training revolves around that other flavour of the month – 'deep learning' – with such systems being able to piece data together and present it to the user in a form that almost suggests that there is a human assembling the "answer' to the specific query. In this sense, such systems are being promoted as tools for enabling a wide variety of verbal and written media to be produced 'automatically', everything from student essays to somewhat more 'profound' PhD theses, serious newscasts, indeed anything that requires textual communications. Picture- and sound-based systems of course are on their way.

The big question of course is 'How good are they?' So to test the software, I like the many other millions who recently signed up to OpenAI (https://chat.openai.com/chat) wanted to figure out how the system could help me improve my search for facts unknown (to myself). Let me repeat the example that I used to explore the system for therein lies its value or otherwise. In the last issue of this journal, we published a translation of Felix Auerbach's paper "The Law of Population Concentration" originally printed in 1913 (Auerbach, 2023). This paper anticipated the introduction of what came to be called "Zipf's Law" published in various places from the 1930s to the late 1940s by George Kingsley Zipf (1949). As many readers of this journal may know, Zipf argued that the size of any city is proportional to the inverse of the rank of that city: the second city has a population one half of the first, the third a population one third of the first, the fourth one quarter of the first, and so on …… down the hierarchy or rank size list. One of the things that has intrigued me about the Auerbach-Zipf

debate is 'who first coined the term "Zipf's law"?' especially as it was anticipated not only by Auerbach but also by scholars such as Pareto, Lotka and others (see Cristelli, Batty and Pietronero, 2012). By the time I was a graduate student in the 1960s, the term appeared to have become the established way of referring to this simple scaling of city sizes and many other power law distributions such as income. But Zipf himself clearly did not coin the term or at least he does not refer to any of his own work in his 1949 book using this term. I am guessing but it probably emerged more or less spontaneously after his book appeared.

So I decided to ask *ChatPGT* the question "Who was the first person to refer to Zipf's Law?" not having any idea how the system would respond. Imagine my surprise when the system came back with a couple of paragraphs of information which at first appeared to be quite authoritative. In essence, it said that in a paper by the mathematician who worked with scaling and fractals, Benoit Mandelbrot, he (Benoit) referred to the fact that when he was at the University of Chicago in the late 1940s and early 1950s, an economist colleague in Chicago, George Stigler, referred to Zipf's Law. This seemed quite remarkable to me that *ChatPGT* had gone off to its archives and dredged up this set of facts for it also told me the paper where Mandelbot had made the reference although it did not have the actual date. It seemed too good to be true. So I decided to check it out. In fact, the paper by Mandelbrot which I got from JStor did not refer to Zipf, so I thought ChatGPT had got the date wrong because back more than 50 years ago, the same papers were often published in different forms but in vain, I could not find anything. George Stigler himself won the Nobel Prize for Economics in 1982 for contributions to the economics of information and it seemed to me entirely likely that he had referred to power laws and scaling. In fact there are plenty of references to his work online but nothing as far as I could see relating to Zipf.

In fact, Wikipedia notes that his son is an eminent statistician, so I decided to email him, also at the University of Chicago, but he could not remember his father ever having referred to Zipf's work in this way. In fact, if you key in questions which directly ask if George Stigler ever referred to Zipf, one draws a blank and if you ask the question, did Mandelbrot refer to Stigler, another blank. Its seems that for the questions I was asking, ChatGPT gave ambiguous and incorrect answers if you moved the elements of the question around; answers that at first sight appeared plausible at second sight looked increasingly dodgy. If you continue by cross referencing Mandelbrot, Stigler and Zipf with others involved in power laws and scaling, you are forced to conclude that there is no clear answer to who first coined the law. The reference to Stigler seems like a dead end. The picture painted by *ChatPGT* is thus confused and distinctly unhelpful.

It is clear that the way one phrases the question makes a difference to the answer but assuming the questions asked do not imply trickery or fake data or opinions, then questions to *ChatGPT* like "Who coined Zipf's Law" should provide fairly robust answers. They do not. This does not prove unequivocally that *ChatPGT* is of no use. Such systems can be used quite creatively and if you consider search as something that is a to-ing and fro-ing of ideas, it can be quite useful. You might object to my example which perhaps lacks context but any query is always limited in this way, and if the system produces incorrect answers, this reveals that in these kinds of weak AI system, then the degree to which these systems are flawed in key ways needs to be clarified, exposed even before they begin to be widely used. Of course there is a vast amount of data out there on the internet and a lot of it might be flawed, so methods

which involve constructing good models of how we might access and use such information are ever more urgent. In fact *ChatPGT* is only one of several systems being developed at present (*The Economist*, 2023). Google has their own called Bard and Baidu have Ernie. OpenAI is part owned by Microsoft while doubtless Amazon and other big platform companies have their own systems in production which are all designed to improve search.

To an extent, focusing on Chatbots, opens the door to the real dilemmas associated with AI. The current perspective in AI is on machine learning where 'explanations' are generated through relating plausible factors that determine how systems function and behave through networks of relations whose structure is iteratively evolved to produce the best predictions possible. Neural networks and like relational structures are used to determine the way data is synthesized to generate good predictions and it follows that the deeper the layers of these relationships are built, then the more likely that useful outcomes are produced. However deep learning goes nowhere if the factors that are used to determine explanations are nonsense in the first place. Where deep learning has been most successful is in contexts where the problem is highly routinised such as games, where the data is enormous in extent and where basic patterns which optimise outcomes are generated by successive manipulation of the training data using the most powerful computers available. The way in which basic patterns that determine good outcomes are explained using such relational structures invariably does not mirror the way we as human beings might arrive at the best outcomes. In fact, some hail the success of this form of AI as being the ability to find new patterns that are associated with the best outcomes that we as humans would never be able to generate by ourselves anyway. The essence of artificial intelligences such as those developed by the company DeepMind such as AlphaGo and AlphaFold fall into this category.

Behind *ChatGPT* there are associations that bear little relationship to the way we might search and relate data and some of the manifest incorrectness of the answers generated to perfectly normal queries clearly relates to the fact that the machine learning use to generate answers is inevitably limited. In the early days of AI, it was argued by the pioneers that more causally determined decision-making models which mimicked the way we ourselves as humans came to good decisions was the way forward but the task proved too great and the field then developed learning methods which sought patterns in data that were no longer related to optimal human decision-making (Wooldridge, 2020). This then is the dilemma faced by new methods of search that attempt to relate data in the manner that *ChatGPT* is designed for. The logics involved are not causal in any sense but statistical and for many of the questions that such system are designed to answer, the most human of characteristics are required. How they can be embodied in the training data is almost an impossible tasks to achieve. Many current applications of AI do not require solutions that we as humans might derive for the focus is on deriving different, usually better solutions, or solutions that we would have no chance of deriving ourselves.

Chatbots like *ChatGPT* thus have little chance of generating solutions that require the sorts of intuition that we as humans use in associating rather unlike things. Chatbots in fact may be able to do this but the unlike things are not likely to make much sense, just as my test about the origin of Zipf's Law did not yield answers that were correct. As *The Economist* (2023) reported, Google withheld its own Chatbot – Bard – from the market, largely because of difficulties over the correctness of its outputs. It was finally released a month ago on 7th

February 2023. One of the things we need to live with in the development of AI technologies is the fact that to make real progress, we need to combine AI with our own knowledge and intuitions. To an extent, this is quite widely accepted in the AI community but the hype that now besets the development of the field and its public image ignore this central message that good AI can only come from merging these systems with our own exercise of logic in human thinking. This is a message that needs to be emphasized and reinforced time and again as we develop more and more systems that combine data in new ways that we find hard to understand.

Michael Batty
University College London

References

Auerbach F (2023) The law of population concentration. *Environment and Planning B: Urban Analytics and City Science* 50: 290-298, translated by Antonio Ciccone.

Batty M (2005) Icons of the early 21st century: 1. Storage and search. *Environment and Planning B 32*: 161-162.

Batty M (2006) Icons of the early 21st century: 2. Security and surveillance *Environment and Planning B 33*: 1-2.

Cristelli M, Batty M, and Pietronero L (2012) There is more than a power law in Zipf. *Scientific Reports 2* 812 8 November: 1-7.

The Economist (2023) The Battle for search: What do AI chatbots mean for the lucrative business of searching the Internet, & Seeking change *The Economist* February 11 p. 9 and 57-59.

Wooldridge M (2020) *The Road to Conscious Machines: The Story of AI*. London: Pelican Books.

Zipf G K (1949) *Human Behavior and the Principle of Least Effort* Reading MA: Addison-Wesley.