

BORGES: A VINAIGRETTE

ROGER KOENKER

In the novel *Galatea 2.0* Rick Powers imagined a proto-LLM trained by the narrator and a CS colleague to pass a qualifying exam in English literature. We seem to be much closer to the reality of this scenario today than we were in 1995 when Powers’s novel appeared, but we still have a long ways to go.

Of course, Borges was there well before Powers and judging from a little googling, there has been quite an outpouring of Borgesian commentary about the current state of AI lately. I was aware that Walter Sosa-Escudero, an outstanding econometrics student at UIUC, now at Universidad de San Andres, wrote *Borges, Big Data y Yo*, several years ago. Unfortunately, it is still not translated from Spanish.

Just last week, Léon Bottou and Bernhard Schölkopf arXived a piece called *Borges and AI*, that explores connections between contemporary LLMs and the work of Borges, principally through *Garden of the Forking Paths*, and *Library of Babel*. The latter story seems particularly apt since the Library includes not only the known corpus of world libraries, but all the false catalogues comprising permutations of the extant corpus. The task of the librarians is to find vindications, “books of apologia and prophecies that would vindicate for all time the actions of every person in the universe.” This sifting might be thought to be the job of science, or even philosophy, not something that could be outsourced to mere machines, but such sifting is also the aim of the Chatbot as it appends token after token to produce a response.¹

Vindication is hard, there are many ways to continue a story, but only a few good ways. This was brought home to me while reading about the recent Gino/Ariely controversy in behavioral economics. If we were to judge Harvard Business Review articles by the standards of the fiction contributions to the *New Yorker* or the *Sewanee Review*, surely these controversial papers would never have appeared in print. They begin with a small psychological conceit to be explored, some

¹My experience with Geppetto has been limited to a few queries about sed/awk constructions, for which it seems remarkably well prepared. For a less successful experience, I recommend Elif Batuman’s hilarious account of asking Geppetto to find a passage from Proust. <https://eliflife.substack.com/p/my-conversation-with-chat-gpt-about>

characters – usually undergrads – who never appear as more than one dimensional just data points in a misaligned excel sheet. There is no plot, or perceptible philosophical or political point. Gross abuse of data seems the least of their sins. On the other hand, when there is so little other merit in the enterprise maybe respect for the data is the least we should expect.

On a tangentially related topic the estimable David Donoho has recently arXived a paper on “Data Science at the Singularity”, a meditation on the cataclysmic consequences of AI that may lurk just around the corner. He identifies three factors that have contributed to the rapid recent progress in LLMs: Data availability and sharing, ease of code sharing, and commonly agreed upon challenges with accepted measures of performance. The first two are familiar – they are just the reproducible research agenda that many of us have pushed for a long time. The third is new, although Donoho’s Tukey lecture already emphasised it in his discussion of machine translation. I found the whole argument very compelling, but the message seems a bit discouraging for economics. Economics lacks a common objective. Unlike computer vision, or machine translation there is repository of tasks whose performance can be judged by a common standard.² Maybe this is what will save us, while auditors, financial managers and lawyers are eaten alive by future generations of Gepetto. The more speculative part of the Donoho paper involves whether AI will evolve into a few massive hegemons, or fragment into many smaller units driven by innovations in software and hardware. For the moment his “brutal scaling” scenario seems to holding sway as start-ups are consumed by Amazon, Google and Meta, but maybe there is hope for a turn-around. There is a curious tension in Donoho’s three factors all of which thrive in the chaos of the open source arena, and the tendency to produce hegemons of enormous power due to sheer computational capacity.

²Like performance on the suite of Atari games for reinforcement learning, Bellmare, et al (2013) *J. Artif. Int. Res.*