

The Construct Editor

Tweaking with Jane, Writing with
Ted, Editing with an AI?

Caroline Bassett

ABSTRACT

New AI language modelers are increasingly capable of taking on multiple editing roles, previously thought to demand humans. Where are the limits? What does further automation of the scholarly editor's function suggest? Will the Construct Editors of the (possible) future render today's scholarly editors redundant? A thought experiment along these lines suggests that editing tools are becoming editing agents — a future of tighter collaboration rather than replacement appears to be likely.

“IT HAS BEEN 25 YEARS SINCE A REPORT OF ORIGINAL RESEARCH was last submitted [. . .] FOR publication”; in a Ted Chiang short story, human science has become a matter of attempting to follow the intelligent machines.¹ Academic journal articles are now exercises in translation as human scientists engage with machine-level thinking that they cannot fully grasp.

What would happen to humanities research in these circumstances? Would humanities work include attempts at exegesis of the literature machines generated for themselves? In the 20th century the literary programmer and OuLiPo member Raymond Queneau argued that only a machine can read a sonnet written by another machine — and his own automatic sonnet *Cent mille milliards de poèmes*, which is by virtue of its inhuman scale impossible to close-read in its entirety, might suggest why.

Setting machine-originated literature aside, what would happen to other humanities activities, for instance to translation work, bibliographic research, scholarly editing? All those activities demand decisions based on multiple materials, on assessments of provenance, on copies and originals, on texts that are partly lost in time, or over time.

1. See CHIANG 2015, 237.

What would be left for the textualists, the close readers, the scholarly editors, and the translators-by-hand to do if human materials were efficiently collected, read, sorted, organized, assessed, compared and annotated, even translated, by AIs? Moreover, how would it be possible (for humans) to judge the outputs of AI processes based not on meaning but pattern recognition, or to question decisions made by AIs based on logical calculation not possible to fully unravel, alien in their rationales and methods? Will we (too) be reduced to tweaking what is already substantially complete, or attempting to account for decisions made by intelligent machines whose workings are unavailable to us?

If these questions seem pertinent today — in life as well as in fiction — this is because recent and rather rapid advances in machine learning and language processing, particularly via transformer models such as GPT2 and 3, have expanded the capacity of AIs to deal in and with natural language. What was a few years ago discussed as pure speculation is now claimed to have some real purchase, even to be something already possible *almost now*. The adoption of a tense beloved of technological futurologists, who are notoriously wrong about the timing of the revolutions they pronounce, is intended to point out this *remains* speculation; but still, something has changed. Beatrice Fazi sees this in terms of a reversal; what began as an attempt to find ways (or develop programming languages) to enable computers to translate human operations *into* readable code, so that the translation trajectory was *into* the machine, now becomes a matter of translating what comes *out*. Debates around machine learning and explainability — or how to deal ethically with the outcomes of black-boxed computational operations, undertaken at scales humans cannot comprehend and that are using logics that are inaccessible or alien — respond to this problematic.

One way to explore these developments is to focus on what is likely to be produced as an output. In the field of scholarly editing, this demands asking what a fully machine-determined edition might look like, how it might differ from what a human editor would produce, and how it might be weighed up or judged. It also means considering how fully AI-informed productions might be different from those *already* produced through extant human machine co-operation; some form of hybrid production is, after all, already more or less standard today, humans having long since integrated digital tools into scholarly activities in the humanities. The trouble is, of course, that as we have already acknowledged, weighing up these productions is not easy; humanities scholars may not seek proof in their work, but there is an expectation, stymied here, that they can account for what they

produce. Argumentation and citation are both, after all, ways of showing the editorial work process. So, let's shift from considering a change in the output to exploring a change in the editor.

"My mother was a computer", said the theorist N. Katherine Hayles, not claiming cyborg ancestry but rather pointing out that in the mid-twentieth century a computer was a human operator. It was only later that the term came to designate a machine. "*I am an editor*", say many redoubtable academics. But will you be, soon? If AI Editors can make a better-than-good guess; a best guess, even *the* best guess — about what "should" come next, about how a text might be completed, or what it should "be" — are they thereby challenging, by suggesting ways to fully automate and therefore render obsolete, the human scholarly editor?²

"*I am an Editor*", says the AI. This is to anthropomorphize, although to be clear I am not arguing that machine learning AI language processors are becoming "like humans" even if they *are* coming to deal in their language in more advanced ways. Letting the AI Editor become a person for a moment is helpful because it demonstrates the shift under discussion, something not to be summed up as amounting to a simple advance in editing tools, nor to an expansion in data available, but centrally involving a shift in the locus and form of decision-making. The AI Editor has agency, a capacity to take decisions (if/then . . .), it replaces the cognitive labor that designates the editor as editor. Such an editor might in the future be able to read into a work, discern its specificities, characterize its typical workings, contextualize its style in relation to genre, compare versions, seek fakes, expose inaccuracies, find new concurrence, make decisions. This vision says that the AI becomes the scholar. The scholar meanwhile becomes redundant.

One day I met my artificial aunt. Anthropomorphism can be taken further. Why limit things to automating the scholarly editor if you can — through the magic of technology — revive the writer "themselves" and force them to speak? A company developing ways to exploit AI language models plays with this possibility, giving the public the chance to ask advice of the dead: "*If I were writing in the 20th century I would have Emma promote her self-published book on Facebook*", so says "Jane Austen", an AI construct of a kind, the artificial progeny of a writer who famously did not marry and so had no issue, although she did have nieces and oddly enough I am one of them.

2. In so far as language AIs plagiarize language itself, they are cheats on the scale of the symbolic.

Of course, there is no Jane, nor is there a Mary (Shelley), although she also has been made to speak on the same site.³ Nevertheless, and although the technology is ragged and in its infancy and the digital afterlives produced gestural or skeletal, somehow the whole is suggestive in relation to scholarly editing and its future trajectories. Could a future “Jane” be put to work, employed as editor of her own edition, perhaps routing around previously necessary intermediaries, those scholarly editors, altogether? The paradoxical promise here — paradoxical given the artifice involved — is for a form of unmediated contact, one that might, by virtue of this apparent collapse of distance, be regarded as authoritative. Revenant Jane and her digital afterlife have a gestural kind of being so far, but would such a construct, even far more developed, ever be the “best” “assessor” of their “own” “original” “intent” anyway? I invoke the scare quotes here to point to what might be lost in the replacement of hermeneutic with quantitatively derived values.

The dream of direct connection, to one another, to an author and their intent, or even to ancestors, that is expressed in the invitation to talk to the celebrity dead is a familiar one, haunting technological culture, not least because it ghosts out the technology that is central to its operations. Another widely held view ghosts out humans.

“*One day I am studying math and notice I have a new key. It is the key to life*”, so an AI told me.⁴ But the language AIs study is not math and not life but human discourse. GPT models work with giant data lakes (copies or parts of the internet) and are then tweaked on smaller datasets. As a trained “authority” on a period or genre, or area, they are set to work. (Perhaps in this sense the AI language generator is always-already a scholarly editor, with a wide grasp of a “textual tradition”?) What machine learning algorithms learn thus includes something human from the start; these intelligent agents, apparently autonomous, often claimed as such, are never “free” of human influence and nor, therefore, of bias (which becomes algorithmically amplified on social media as is now widely recognized). They are suspended in a giant human soup, a discourse containing multiple times, periods, actors; ingredients tasted simultaneously, without restraint.

-
3. The invitation to talk with the dead is egregious, of course, in some situations. In relation to archives of dispossessed peoples for instance, attempting forms of dis-interment — making the dead speak — may be unethical (see e.g., ODOMUSO 2020).
 4. Automatic text produced using the methods described in the Cambridge Digital Humanities methods workbook, *Text generators: from DADA to GPT-3* <https://tinyurl.com/textgenworkshop>.

What do they create with all this? Nothing says computer science. These algorithms do not create, they simulate. They produce fake news, simulate human agents, or on another register simulate language itself — the words they produce appear to make sense and therefore appear to be informed by a logic that relates to meaning-making but is not. These are stochastic parrots, Timnut Gebru and her co-writers insist. They can produce nothing new under this sun — not least because they have no experience of it. This doesn't mean that language AIs cannot act. They have a form of cognitive agency, a growing capacity to work in language and to work upon texts;⁵ it might mean their editorship is mechanic, that the editor function decays.

So what have construct AIs, dead parrots, and speculative fiction have to do with scholarly editing and its potential automation? What are its possible futures in a highly computational culture? If we dispose of dreams of the Machine Editor with absolute authority (all the data, all absolutely un-marked by human factors; the dream of something outside of culture), if we reject even as a fantasy the idea that a simulated personality brought back to pronounce on their own work could do an adequate job — not only because the revenant construct author would not be the best judge of their own work outside of the contexts and times of its immediate production,⁶ but also because the construct fiction obscures the mediating role of the algorithm — and if we admit the justice of the arguments that these agents, by virtue of being AIs, do not create but simulate, then what is left? Are the new digital editors simply going to be tools for humans like those that came before them — just slightly better at doing what earlier generations did?

It seems to me that what is required to bring some clarity here is, first, recognition of the growing agency of AI-based algorithms, which will produce editors that are not quite like the tools that came before them, and, second, recognition that this agency does not make these AI agents autonomous, just as their capacity to produce does not make them creative. The key might be that they do produce — or rather as we have set out to indicate — they co-produce. That is, they are entailed in new kinds of partnerships with human scholars, albeit not ones in which each partner can assume they will ever fully understand the other. It is with this in mind

5. The internet is full of words and writing, notes Richard Seymour in *Twittering Machines*; more and more and more of it is generated by AIs.

6. Christopher Ohge notes that for many (human) scholarly editors it is axiomatic that an edition does not intend to gather up the “author’s purified intentions” but rather to set them in the myriad social contexts of a work’s production and dissemination (see OHGE 2021, Chapter 2).

that it occurs to me that translation, seen by Ted Chiang as a reduction (what is left for scholars locked out of knowledge generation), might be invoked otherwise: as a way to think about what might be produced, as an on-going process entailing increasingly complex interactions between human and machine “editors”; and if one of the collaborating “editors” here is a scholar, the other is not an idiot.

University of Cambridge

Works Cited

- ALEXANDER, Anne, Caroline BASSETT, Alan BLACKWELL, and Jo Lindsay WALTON. 2021. *Ghosts, Robots, Automatic Writing: An AI Level Study Guide*. Cambridge: CDH.
- BENDER, Emily M., Timnit GEBRU, Angelina McMILLAN-MAJOR, Shmargaret SHMITCHELL. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”. *Proceedings of the 2021 ACM Conference on Fairness*. <https://doi.org/10.1145/3442188.3445922>.
- CHIANG, Ted. 2015. “The Evolution of Human Science”. *Stories of your Life and Others*. London: Picador.
- FAZI, Beatrice. [2018] 2019. “The Discrete and the Continuous”. *Theory, Culture & Society*, 36.1: 3–26.
- HAYLES, N. Katherine. 2005. *My Mother Was a Computer: Digital Subjects and Literary Texts*. Chicago: University of Chicago Press.
- HERN, Alex. 2019. “New AI fake text generator may be too dangerous to release, say creators”. *The Guardian*. 14 February.
- ODUMOSU, Temi. 2020. “The Crying Child: On Colonial Archives, Digitization, and Ethics of Care in the Cultural Commons”. *Current Anthropology*, 61.22: 289–302.
- OHGE, Christopher. 2021. *Publishing Scholarly Editions: Archives, Computing, and Experience*. Cambridge: Cambridge University Press.
- QUENEAU, Raymond. 1961. *Cent Mille Milliard de Poèmes*. Paris: Gallimard.
- SEYMOUR, Richard. 2019. *The Twittering Machine*. London: Indigo Press.