

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

"In the wake of Ovid's unicorn"

Permalink

<https://escholarship.org/uc/item/65z3479m>

Author

Raley, Rita

Publication Date

2019-11-01

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed

Rita Raley, “In the Wake of Ovid’s Unicorn” (November 2019)

Presentation for the annual conference, Society for Literature, Science, and the Arts (UC Irvine)

<https://litsciarts.org/slsa19/>

// This script is the presentation as delivered, with verbal elaboration of the “10 points.”

Human-generated prompt: How should we teach natural language processing, which is to say human reading and writing, after the success of large language models? What should be the response of language instructors to the exponential developments in the field of neural text generation?

First try, from GPT-2: This talk explores those questions and looks at the state of the art of neural text generation and the prospects for future work.

As directed: Nine months ago, OpenAI announced its general language model, GPT-2, but did not fully release it. Trained on 40GB of English-language data, with 1.5 billion parameters, the model was said to perform almost too well; a risk analysis concluded that the team could not be confident that potentially negative consequences would outweigh the positive and advised early caution and limited release. (For comparison, Google’s BERT, released three months prior, had 345m parameters; I should say that I am focusing on OpenAI because it tells a story.) **{SLIDE}** The model’s capability then could only be assessed by the general public through “synthetic text samples” of which this was the most famous: “Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.”¹ Absent meaningful and comprehensive metrics for evaluating the quality of the output of any language model, I can only offer my subjective opinion and say that this works. **{SLIDE}** Why it does is somewhat surprising, as researchers from the University of Washington observed in their analysis of the small GPT-2 model, which had only a fraction of the parameters. What is particularly counter-intuitive is that the highest quality samples resulted from a degree of randomness rather than *maximum likelihood*, as one would expect to be the case for predictive text. Adhering to rules and patterns is a common strategy of maximal probability—so the less probable the move, the greater the surprise. The naturalness of the prose in this sample, then, the uncanny liveliness of its reporting on the discovery of a herd of unicorns in the Andes mountains, results not only from its adherence to a journalistic template, but also from its slight break from the obvious and the expected: a “natural fountain” rather than a waterfall, and the naming of the population, Ovid’s unicorn. One conclusion to draw is this: humans may seem to display a preference for appropriation, mimesis, and memetic expression—everyone is always copying everyone else—but in actual linguistic practice turbulent distribution is the mark of an authentic “human” style.

{SLIDE} Three months ago, two graduate students replicated the 1.5b parameter model (as did others apparently); and OpenAI soon thereafter released the medium model of GPT-2 (774m parameters). **{SLIDE}** This week, they did a full release, citing an only marginally better credibility score assigned to its output, and now it is possible for everyone to play with the whole thing. So for what it’s worth, if one talks to the transformer, as the online implementation invites, it is indeed possible to verify the claim that this is a “better model.” **{SLIDE}** Using the same prompt, I got this text after two tries: “A team from the University of Edinburgh spent months collecting data about the unicorns’ behaviour, ecology and diet. They also found that the unicorns were much more social than their relatives. In addition to their amazing language skills,

the researchers also discovered that the unicorns are much bigger than the average cattle. ‘They’re a giant-sized sheep that have taken on a mammoth-sized body,’ said Dr. Michael J. McManus, head of the research team. The researchers have named the unicorns Santicus and Santicus de Luna, which are Spanish and Latin for ‘small unicorns.’” It can be said again: GPT-2 works. Voice, concrete details, backstory—all of these help facilitate quality output but it is clear why OpenAI would again have cautioned against misuse and noted that the “detection of synthetic text is a long-term challenge.” **{SLIDE}** What goes to my knowledge unremarked is the question of GPT-3; if the model is continually scaled up and more data added, at what point does the behavior qualitatively change. But truly we are already in a new phase, and that is the era or age of numerical rather than lexical representation.

To amplify: **{SLIDE}** to produce his novel, *A Noise Such as a Man Might Make*, which appears in the Nick Montfort’s new Counterpath book series, “Using Electricity,” Milton Läufer combined Ernest Hemingway’s *Old Man and the Sea* and Cormac McCarthy’s *The Road*. The pairing works for perhaps obvious reasons: they both involve struggle against hostile environments and thematize masculinity, survival, and perseverance; there are no proper names and no points of resolution. Apart from syntactic similarity, the texts seem to share a style and tone—features that cannot necessarily be concretely identified in a text but are rather something one has a “feel” for. Läufer tokenized the two source texts, creating conditional frequency distributions for bigrams, trigrams, and tetragrams, and the resulting synthesis ends up reinforcing or making concrete that intuitive insight. ‘I think they’re alike’ thus becomes ‘look, they are alike.’ (He probably did some cleaning but the final document is coherent.) What I wish to suggest with all of this is that there is a meaningful, even perhaps epistemic, distinction between generating text from tokenized documents on the one hand and word embeddings on the other—that is, vectorization, translating words into numbers that can be read by machine learning algorithms.

But if it can be said that we are in fact living through a break, insofar as it can be understood to be evental, it is not only because Natural Language Processing is using pretrained language models and thus having its watershed “ImageNet moment” (‘P’ in GPT is pretrained). **{SLIDE}** I think any account of a break must also include the development of WebText **{SLIDE}**, the corpus that OpenAI created to train its model. If prior training data from Project Gutenberg and Wikipedia tacitly suggested “the common word exact without vulgarity,” which is to say standard English, with all the notions of the proper and the correct that implies, a corpus of some 8 million webpages scraped from Reddit posts with a minimal number of karma points suggests instead that there is no common word, training for a language model that does not itself model communication. For language models at their current scale – NVIDIA recently announced that it was training a transformer language model 5.6x the size of GPT-2 – Gutenberg and Wikipedia are now too small, delimited, and paradoxically singular, their relationship to language too proprietary and protocological. WebText, by contrast, eliminates any trace of source text, or buries it rather, and results in non-indexical output, language that does not point back to a discrete place of origin. The work of authentication, then, is displaced onto the upvotes of the crowd, understood as a “heuristic indicator for whether other users found the link interesting, educational, or just funny,” although we might ask where such properties are to be located: in the user’s framing of the link—which is to say, in their interpretative activity—or in the external reference. Is the link the implicit locus of expertise, in other words, or is it the activity of recommendation itself?

We know where GPT-2 leads us: **{SLIDE}** to a “small valley” in the Andes Mountains, with unicorns but “no other animals or humans,” a fantastic place invented by a machine, and uninhabited by people. It leads us then, to the portending of our own obsolescence, to a future in which the world has no need of our writing, and to a world that might itself be uninhabitable. So, in the wake of Ovid’s unicorn, how might it be necessary to change how we teach literature and the language arts?

Or, framed as another human-generated system prompt **{SLIDE}**: how should pedagogy adapt and respond to the exponential developments in the field of natural language generation? Do we even have a good understanding of the limits of the human ability to generate and understand the meaning of natural language? How does the field of natural language generation (NLG) inform our pedagogical practice?

[First try, from GPT-2] These are all serious questions, and they need to be addressed. The answers, however, should not just be determined by theoretical research but should also be driven by concrete empirical data from real applications in the real world. We can’t just study how to solve a theoretical problem. We need to actually learn to solve the problems that we face.

{SLIDE} To that end, I have given some consideration to what in pedagogic practice we ought to do, to what should be done with curricula in the language arts—responding not to a threat assessment from the perspective of the institution, but as an expression of a genuine ethical and political commitment to the idea that pedagogy ought itself to be responsive, and responsible to, the fundamental transformations in our linguistic milieu. What then should the training of students entail? 10 points.

1. templates, protocols, genres; more explicit focus on form; how do you recognize dramatic monologue or trochaic meter when you see it
2. rhetorical figures
3. mimesis (writing like, recognizing and inhabiting voices not one’s own). Vestiges of Wordsworth, the feeling of Lovecraft, that snippet sounds like Kerouac—for a literary scholar, a work that incorporates, instantiates, or is otherwise in-formed by literary source texts can be experienced as a kind of abstract commonplace book, with echoes, intonations, even traces of texts that have come before, but without direct reference to them.
4. collaboration: self-reflexivity about tool use, and about “human” writing as a historical construction. David (Jhave) Johnston’s “custom contemporary poetry corpus,” which he trained for his project, *ReRite*. But it does not have to be that extensive; creative exercises with Smart Compose and online talk transformer.
5. courses in histories of writing as a technology
6. demystify and disambiguate the alphabet as a cultural technique; to that end, consider asemic and a-grammatical, a-grammatological practices; perhaps here new life for aesthetics of the glitch
7. mandate second language instruction; direct experience of being not-at-home in language

8. learn how commands and operations work in programming languages, what is a function in Python, what is the basic syntactical and/or logical structure of a given script

9. long-term discourse on writing as a practice of care of the self: cultivation of psychic life, self-improvement (memoir, journaling, life writing); expressivity; Allison Parrish, “intentional writing”

10. care of the community: detection of disinformation; learn how to verify (not just a literacy problem, also forensic, political); close reading, textual analysis for anomaly detection; not symptomatic analysis necessarily because quite explicit political motivations rather than unconscious ideological commitments, as in faux Bernie supporters or BLM activists)

{SLIDE} Ross Goodwin proclaimed of his work with Benjamin, the LTSM-RNN scriptwriter that, “The camera set painting free. One of the things we’re doing is setting writing free.”² Natural Language Generation, like the camera, has indeed started to become the status quo for basic tasks of documentation and communication. But so too has it been set free, and does its best work—and this has been empirically assessed—not imitating human language, but actually reproducing its function, that is to say, generating speech for people who do not exist. **{SLIDE}** Dr. Perez says, “By the time we reached the top of one peak, the water looked blue, with some crystals on top.” **{SLIDE}** To bring this back to number 10, and the pressing problem of disinformation: truly human-like or human-passable language comes in the form of quotation. As Allen Institute researchers determined from their work on the Grover language model, what particularly makes a fake news story credible, one for example about the link between vaccines and autism, is not the narrative or the presentation of data, but the emulation of expert commentary.

Ours is a profession in transition in more ways than we perhaps realize. What I have tried to outline in this talk is a preliminary way of thinking through how we might begin to respond to these changes that are upon us now, in the wake of Ovid’s unicorn.

¹ <https://openai.com/blog/better-language-models/>

² <https://arstechnica.com/the-multiverse/2017/04/an-ai-wrote-all-of-david-hasselhoffs-lines-in-this-demented-short-film>