345M-GPT-2 After James Wright: Can AI Generate Convincing Contemporary Poetry?

Jonah Zitelli

Kenyon College

INTRODUCTION

In recent years, the ability of AI technology to generate creative content has progressed in leaps and bounds, especially in the realms of visual art and music. However, even the most advanced modern language models, such as GPT-2, AI still has not succeeded in replicating the as of yet uniquely human propensity for creating literary texts. Poetry is widely considered to be one of the more difficult literary forms to master even for humans, as it is by definition more sparse, and therefore denser than prose writing—Poetry distills emotion with the aid of metaphor, imagery, and more formal considerations such as line breaks and rhyme schemes. So, while GPT-2 has had some measured success in modeling screenplays and the work of playwrights, it would not be unreasonable to expect it to struggle to generate coherent poems.It has been previously shown to understand meter and rhyme with enough training input, but has fallen short with regard to coherence [1]. With that in mind: How well, after being trained to replicate the poetic style of mid-century Ohio poet James Wright, can GPT-2 put that style into practice while generating new works? Where are its strengths and shortcomings?

BACKGROUND ON GPT-2 TECHNOLOGY

GPT-2, created by OpenAI, is a transformer-based language model with 1.5 billion parameters, and it has been trained on 8 million webpages with the goal of accurately predicting the next word in a piece of text, given all the previous words in that text [2]. This goal makes GPT-2 capable of completing various tasks: It can answer questions about the content and comprehension of texts, summarize, and translate. It can also generate synthetic texts in which it mimics the style and content of a given input [2]. In order to mitigate possible malicious use and provide additional time to analyze the implications of releasing GPT-2 1.5BM OpenAI is taking a staged-release approach, with the most powerful version currently available being GPT-2 345M.
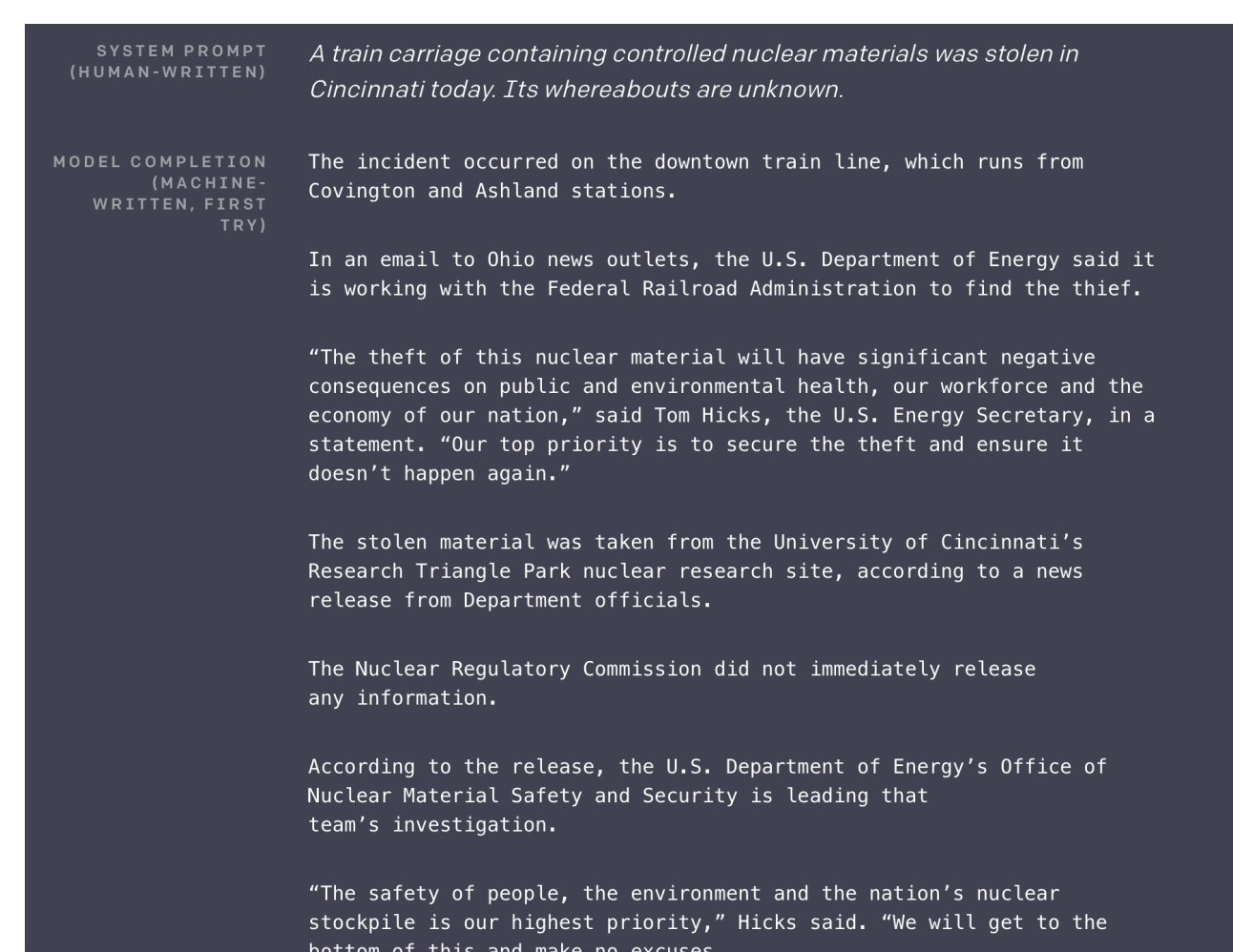
ON THE ART OF POETRY

Poetry is an inherently difficult art form to master, even for human beings. The goal of a poem is to distill language in a particular way in order to achieve a heightened, yet specific emotional effect using relatively sparse, yet heavily descriptive language. As such, poets rely heavily on metaphors and images to relay information in a concise manner while maintaining their desired tone or mood. It is precise, but flexible, too. It is not necessarily bound to standard grammatical structure like most prose text is. Given these aspects of poetry as an art form, it is only natural to expect GPT-2 to have some difficultly generating believable, coherent poetry even when fine tuned to the work of a particular poet or set of poets. 117M-GPT-2 has been shown in the past to have a propensity for generating heavily structured, rhyming, and otherwise "old-fashioned" poetry when trained on the work of poets such as Shakespeare, Tennyson, Pope, and Yeats [3]. However, its potential to generate work with the flexibility of more contemporary poetry, such as that of James Wright, the object of this study, is more doubtful.

METHODS

In order to generate poems in the style of James Wright, a 345M-sized GPT-2 model was fine-tuned to a corpus of 177 of his poems across a span of his entire career. During the training process the text of these 177 poems was fed to the GPT-2 model 6,000 times, with the model's poetic style becoming more like that of Wright each time. Following this, the model used its understanding of the patterns contained in human language combined with its new knowledge of James Wright's poetry in order to generate two sets of poems in the style of Wright. These two sets differ in their temperature, with one being set at 0.7 and the other at 0.9. A higher temperature setting means more randomness will be present in the text generated by the model, so the set of poems with the 0.9 temperature setting should be expected to contain more deviations from Wright's poetic style and content than the set with the 0.7 temperature setting.

RESULTS

Provided here are some of the most interesting samples from the 0.7 and 0.9 temperature sets generated by the 345M-GPT-2 model respectively:

```
Spinning on its side,The boat tumbles down the channel
And, a mile or so further,I can hear trilling in the trees.
I want to leap on its back and protect
My dear friend, the beautiful bird.
Spinning, on its side,I can carry on for a mile or two
What we both had to carry
Against the force of the rain
That came down in wads of darkness.
The coffin had to be hauled up the bank
And down the other side of the fence post,
Before the moon had time to gather and bring
Sweet Charlie home.

Now, a mile down,I can still see the spot where he went
Out of sound asleep and into sound
Dead.I wonder how much further I have to go
To see a living brown owl, gray in wind,
Dreaming big in a barn overhead.
```

```
I stand here in the sight of everyone, I see my own face
Look at your teeth chatter in the air,
And yet you keep silent. You gaze up at the sky
Resonating with the voice of the hunted
Two humps above the murmur of the brooks.
Both of them starved down there,
Beside the open wound where the pony
Strapped to the tree, between the trees.
The other man lay reeling, panting,
Brookside to Brookside, down, down, down, dead.

I take the tree by the humus and lean
Keen for a liar, quick to jump to conclusions,
Strong enough not to trip my spine.
One bell before I step outside
Broke the silence. A tall
====================
In the Ancient City, Lazarus
Walked, alone,Through a Mine of Lake Miners
Where Oilcloth Don Quijote Laredo.
He Cried, And Someone Cleave
Alien And WyReliefThroughThe Rue Ambroise.

And the Call To ArmsFalledLooseAt Last
Down Near The Iberian Islands,
AndVeronica Cole
Renowned As A Poet
Desperate For The Kiss
Of A Fresh MarineBornOnBoard The Rocinante.

Where The BlackbirdSpeaksSweet Song
Wind Beneath A Sand-Hair Dune.

I Need No Water,And Lara LoganIs Dead.

If I Only Had A BlanketTo Float Dead
In, A mailboxShould Do The Trick.
```

ANALYSIS AND CONCLUSIONS

As expected, the 345M-GPT-2 model trained on a corpus of Wright's poetry encountered problems generating believable, coherent poems in his style. However, it did not go entirely without success. The model was most successful, especially in its 0.7 temperature iteration, in generating work that matched the style and form of Wright's earliest work, which was much more beholden to traditional poetic forms and rhetoric than his later work, often to the point of criticism [4]. The model was fairly quick to pick up on elements like rhyme and meter in and its execution of these poetic structures was surprisingly nuanced—there were multiple examples from the 0.7 temperature set of the model using various slant rhymes in interesting ways, something not achieved in the 0.9 temperature set. The first example at left was the strongest replication of Wright's later work generated in the 0.7 temperature poem set. The model was good at picking up on vocabulary content and themes, using these two elements to capture Wright's tone and maintain it more or less consistently. Its core images and phrasing are believable for Wright. It falls short when it comes to keeping images and metaphors consistent over the course of the poem, instead overloading the text with many disconnected images.

The second example at left is especially interesting. It comes from the 0.9 temperature poem set, which contains more randomness, allowing for the grammatical and syntactical flexibility characteristic of much contemporary poetry. This works well in the poem's first half. The model is successful at capturing Wright's distinctly understated phrasing, and combining it with unique images that are nonetheless reminiscent of Wright's work. However, in the second half of the poem, the model becomes mostly incoherent, pulling words from corners of its lexicon unrelated to Wright in the slightest—Note the extremely strange *Don Quixote* references that double as a reference to the novel *The Expanse*, in which "the Rocinante" is a spaceship.

Overall, there is evidence that GPT-2 has the potential to accurately replicate contemporary poems if trained well. Currently, it's strengths lie in capturing phrasing and tone.For this to happen, the model would need to have enough randomness that it avoids consistent rhyme and meter structures while also remaining on topic. Additionally, the model would need to learn to carry an image or set of images all the way through a poem. In theory, this would lead to narrative coherence, an element that was lacking in these sets.

REFERENCES

https://slatestarcodex.com/2019/03/14/gwerns-ai-generated-poetry/ [1]
https://openai.com/blog/better-language-models/#sample2 [2]
https://www.gwern.net/GPT-2 [3]
https://www.poetryfoundation.org/poets/james-wright [4]